TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI INGEGNERIA BIOMEDICA, ELETTRONICA
E DELLE TELECOMUNICAZIONI

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

MODELING, SIMULATION AND VALIDATION OF THE ELECTRO-THERMAL INTERACTION IN POWER MOSFETS

STEFANO DE FILIPPIS

Il Coordinatore del Corso di Dottorato

Ch.mo Prof. Niccoló RINALDI

Tutori

Ch.mo Prof. Andrea IRACE
Dr. Michael NELHIEBEL
Dr. Vladimír KOŠEL

A. A. 2011–2012

Acknowledgments

The research activity presented in this dissertation has been carried out from March 2010 until March 2013 within a successful cooperation established between the Dipartimento di Ingengneria Biomedica Elettronica e dell Telecomunicazioni (DIBET) from Università degli Studi di Napoli "Federico II", the KAI - Kompetenzzentrum für Automobil- und Industrie-Elektronik GmbH and Infineon Technologies Austria. Therefore, first of all, my most sincere acknowledgments go to Prof. Andrea Irace, Dr. Michael Glavanovics from KAI and Mr. Markus Ladurner from Infineon Technologies Austria, who gave me the opportunity to start this PhD research activity. For the same reason, I am grateful to KAI's former director Prof. Herbert Grünbacher and to its current director Mr. Josef Fugger.

I express my true gratitude to my academic mentor Prof. Andrea Irace and KAI supervisors Dr. Michael Nelhiebel and Dr. Vladimír Košel, who wisely guided me throughout the entire research studies with knowledgeable supports and their always motivating, useful and interesting discussions.

Very special thanks go to Dr. Michael Glavanovics. This work would not have been possible without his helpful suggestions during the definition of this research topic. His knowledgeable advice, talks and competent points of view on both theoretical and practical issues of my PhD research have been very important for me. I also would like to express my sincere gratitude to him for helping me to improve the text of this dissertation with many accurate proofreadings.

This research has been conducted in parallel with the PhD work of Dr. Helmut Köck. I thank him for the very pleasant collaboration, constant availability and very stimulating discussions about FEM simulations and other common research topics.

Many thanks go to my KAI colleagues: Gregor Pobegen for very interesting discussions and for his kind help during probe-station measurements; Benjamin Steinwender and Dr. Hans-Peter Kreuter for their competent IT support. Furthermore, I thank my former project manager at KAI Dr. Arno Zechmann.

Last but not least, I would like to thank all my KAI colleagues for the very pleasant working atmosphere.

I thank Mr. Robert Illing from Infineon Technologies Austria for the interesting talks we had concerning both theoretical and experimental aspects of my work.

At Infineon Technologies Germany: I am grateful to Dr. Stefan Decker for stimulating conversations on electro-thermal issues and for furnishing me with TCAD simulation data; Dr. Donald Dibra for providing test chips and advice concerning the measurement setup, and Dr. Liu Chen (Caroline) for interesting discussions on the electro-thermal simulation topic.

Many acknowledgements go to my university colleagues at OPTOLAB. In particular, I would like to express my gratitude to Luca Maresca and Dr. Michele Riccio for their helpful support concerning administration issues.

I dedicate this dissertation to my parents, who persistently motivated, morally and economically supported me throughout my scholar and university studies.
Questa tesi di dottorato è dedicata ai miei genitori, i quali mi hanno sempre tenacemente motivato, supportato moralmente ed economicamente durante i miei studi scolastici ed universitari.

Finally, I express my deepest gratitude to Isi, who has represented and still represents for me a fundamental source of strength. I thank her for everlasting encouragement and support all along my PhD research activity in the more gratifying moments, as well as in the more difficult ones. *Herzlichen Dank!*

Contents

List of Figures

Introduction

Power MOSFETs are massively used in most low and medium voltage power applications thanks to their noteworthy advantages compared to power bipolar transistors, such as their higher switching speed (shorter turn-on and turn-off times) and their larger power capability [1, 2, 3]. They are also increasingly employed in automotive applications (Smart Power switches [4]) to enhance energy efficiency, safety and comfort in latest generation cars. In the last decades they have been subjected to a rapid development in terms of their performance due to more demanding requirements in traditional applications and due to the fast development of new challenging ones. As a consequence, designing higher performance devices leads often to more critical and limiting robustness and reliability issues.

During operation in typical automotive applications a power switch is subjected to critical stress conditions resulting in a high power dissipation which may progressively degrade its electrical performance, shorten its lifetime and, eventually, cause its failure. The high power density often causes very high temperature peaks and large thermal gradients across the power MOSFET. Those thermally extreme conditions are the driving force to fatigue phenomena resulting from mechanical stresses, which are a consequence of the different thermal expansions experienced by materials composing the device [5, 6, 7, 8, 9, 10, 11]. Furthermore, main electrical parameters of a power semiconductor device show a substantial temperature dependence which may determine thermally unstable operating conditions leading to unexpected early-failures even during a single stress pulse [12, 13]. Therefore, main reliability and robustness problems in power MOSFETs are caused by fast temperature transients and can be only counteracted with the development of adequate technologies, geometries and circuit designs aimed at achieving an effective thermal management.

A deep understanding of the heat propagation through the device structure during critical stress conditions is of utmost importance when design-

ing new power MOSFET technologies or when improving already existing ones. Several measurement approaches are available to accurately assess the thermal behavior of a semiconductor device. They are usually based on contact methods, like the employment of temperature sensors (often integrated) [14, 15, 16], or on more efficient contactless methods, such as infrared thermography [17, 18, 19] or laser interferometry [20, 21]. Unfortunately, experimental methods can be used effectively only when power device chips, or as an alternative purposely designed test structures, are available. Hence, during the development phase of brand new power devices, employing simulation tools represents the only method useful for predicting the device thermal behavior and the effectiveness of new design proposals. Thermal simulators are commonly utilized for this purpose [22, 23]. Unfortunately, they are limited to specific device operations in which the electro-thermal interaction does not significantly influence the thermal response of the device. Specifically, those tools are intended to solve the heat equation within the device structure and, hence, they neglect the impact of temperature variations on device electrical parameters and vice-versa.

The electro-thermal interaction takes place in a power device due to the temperature dependence of its main parameters, such as the threshold voltage and the carrier mobility. Such phenomena define thermally unstable operating conditions for the device which limit its reliable operation and, possibly, its robustness characteristics. In the last two decades, reasons and consequences of the electro-thermal instability in power MOSFETs have been widely studied, experimentally verified and documented in literature [12, 13, 24, 25]. On this basis, electro-thermal simulation tools were born to overcome the limitation of purely thermal and purely electrical simulators.

Research activities have dealt with the modeling, the simulation and the experimental verification of the electro-thermal interaction in power MOSFETs. Specifically, the attention has been focused on low-voltage power MOSFETs integrated in modern Smart Power switches and on their electro-thermal behavior in critical operations defined by the harsh automotive environment. In the **first chapter** of this thesis a brief introduction will be given about most common power MOSFET structures. Subsequently, simple physics-based models of the device elementary cell are proposed in order to describe its electro-thermal behavior. This modeling is necessary to accordingly simulate the electro-thermal interaction within a power MOSFET. Finally, some basic notions will be reported about modern Smart Power switches.

An electro-thermal simulator is a software tool capable of predicting, by

applying numerical computation methods, the device electrical response, its resulting thermal behavior and finally to consistently couple their interaction. Generally, two main approaches are utilized to implement such a coupling and they both show advantages and drawbacks. For instance, simulators based on the relaxation method exploit a coupling strategy which requires a laborious implementation of data transfer synchronizations among the electrical and the thermal simulator [26, 27, 28, 29]. On the other hand, simulators using just one single software environment (direct method) avoid synchronization issues, but have to be programmed from scratch implying a high implementation effort [30, 31, 32, 33]. Furthermore, this approach requires an accurate electrical and thermal modeling of the device which is generally not an easy task.

During the research activity, an FEM-based electro-thermal simulator has been developed and implemented by customizing the ANSYS® software by means of its native design language. A novel physics-based concept has been developed for consistently coupling the electrical and the thermal behavior in a power MOSFET by exploiting ANSYS electrical and electro-thermal simulation capabilities [34, 35]. Such an approach avoids the employment of two distinct simulators (which represents the source of main limitations for the relaxation method) and, at the same time, simplifies the device thermal and electrical modeling (main limitation for the direct method). Lastly, the conceived algorithm can be implemented with a reasonable effort also in other commercially available FEM suites, like COMSOL Multiphysics®.

Specifically, two distinct simulation concepts have been developed. The first one is suitable for simulating transient operations in discrete power MOSFETs starting from assigned V_{GS} and V_{DS} biasing conditions. In the second one, the same concept has been extended towards electro-thermal simulations of power MOSFETs integrated in Smart Power devices. In those devices, the gate pin is generally not directly accessible due to the integrated gate drivers. Furthermore, device biasing conditions during critical operations are generally defined by the external circuitry. Therefore, the latter algorithm for the prediction of the thermal and electrical field is based on the assessment of the internal V_{GS} voltage which agrees with the externally load-imposed biasing conditions (i.e. V_{DS} and I_D) [36].

The **second chapter** is devoted to accurately presenting both developed simulation algorithms. Two test devices have been employed to validate simulator accuracy. Specifically, experimental measurements have been performed thanks to temperature sensing structures integrated in the test chips. Therefore, main features of such temperature sensors will be explained in detail together

with their thermal characterizations.

Simulations of power devices establish a typical multi-scale problem in which the large device structure includes micro-structures which are often highly anisotropic, such as the metal interconnection stack or the gate trenches within a power trench MOSFET. The geometrical modeling of multi-scale structures represents a challenging task since a detailed modeling of its geometry is required to achieve accurate results, but it often results in an unacceptable computational effort. In this work, the homogenization approach ([37, 38, 39, 40]) has been used in power devices with the goal to estimate thermally equivalent material properties of micro-scale structures. In particular, this approach has been used for the first time to determine thermally equivalent material properties of the epitaxial layer of a power trench MOSFET [41]. In this case, an accurate equivalent description of the epitaxial layer may be relevant since this represents the device region where most of the power is dissipated due to the Joule effect. In this context, for such an epitaxial layer, a physics-based electrical modeling has been developed as well. The thermal homogenization approach, its utilization in a power trench MOSFET and characterization results will be treated in the **third chapter**. Furthermore, the impact of such innovative modeling on the description of the device electro-thermal behavior has been evaluated and compared to a commonly adopted isotropic model of the epitaxial layer.

In the **fourth chapter**, two applications of the developed simulator have been reported in detail to demonstrate: 1) the importance of the electro-thermal interaction in modern and future power MOSFET technologies; 2) the usefulness of the developed electro-thermal simulation approach.

Due to miniaturization reasons, the trend in modern Smart Power switches aims to achieve MOSFETs featuring smaller active areas. The consequently lower current capability of the device is generally counteracted by developing new technologies with progressively diminished ON-state resistance. This can be implemented by adopting technology improvements which often result in a higher transconductance coefficient K_{MOS}. Electro-thermal simulations of two test devices have demonstrated that the higher the K_{MOS}, the less thermally stable becomes the device behavior and, consequently, the less robust is the power MOSFET. Simulation results have been validated by means of temperature measurements performed using an integrated temperature sensor [36].

Finally, the electro-thermal simulator has been used to estimate the impact of a crack within the copper metallization of a power MOSFET. This kind of

failure mechanism has been observed in modern Smart Power devices subjected to repetitive short circuit pulses [10]. The conducted study has demonstrated that a crack in the power metallization drives the formation of a localized temperature hot-spot which grows higher for increasing crack sizes.

Chapter 1

Power MOSFETs in Smart Power switches

First power devices based on the Metal–Oxide–Semiconductor Field Effect Transistor (MOSFET) structure were developed in the mid-70s to replace the existing power Bipolar Junction Transistor (BJT) in power applications featuring high switching frequency. At that time, power BJTs could not be operated at high frequencies due to their long recovery time required to deplete their drift region from stored charges. Power MOSFETs instead, showed improved switching performance compared to BJTs thanks to their uni-polar nature. This peculiarity represented their most important advantage, however their high input impedance was also a very appreciated feature allowing the design of more simple device drivers. Last, but not least, power MOSFET robustness performance was significantly better when compared to BJTs, especially in the inductive hard-switching behavior [1, 2, 3, 42].

Originally, power MOSFET technologies were based on a double-diffusion process in which the formation of the inversion layer was obtained by controlling the depth of two junctions. Power devices realized in such a way, called Vertical-Diffused MOSFET (VD-MOSFET), had the drawback of a quite limited current capability due to their substantially large internal resistance (also called *ON-resistance* or *ON-state resistance*), which in turn represented a prominent limiting factor for both the device power management and the efficiency of the circuits in which they were utilized. Therefore, in the early 1990s an improved technology was introduced for realizing new generations of power MOSFETs, i.e. the so called trench technology. In general, compared to a VD-MOSFET, a power trench MOSFET shows a reduced ON-resistance

and allows higher operating frequencies (up to several MHz).

In the last decades, along with the progressively improved power MOSFET technologies, automotive applications have impressively grown due to more demanding user requirements for increasing car performance, safety and comfort. Such a trend has led to a great development of automotive electronics towards higher complexity, which has resulted in a growing integration and miniaturization of power switches. Nowadays, most of the discrete switches used in the automotive environment, such as power transistors or electromechanical relays, have been replaced by complex semiconductor-based circuits, the so called Smart Power (SP) switches. Those devices integrate within a single chip one or more power MOSFETs together with several analog circuits for driving and protecting the power devices and the loads, and the logic circuitry for managing the data exchange with the car control unit. In automotive applications, SP devices are mainly used for motor controllers, power supplies and lamps ballasting. In general, a SP device must reliably withstand a wide variety of electrical and thermal stresses during his operation in the harsh automotive environment. Some of the aforementioned stresses result from critical biasing conditions which may drive the power transistor towards a dangerous operating regime called thermal instability. Typically, main MOSFET parameters show a substantial temperature dependence, as a result of that a mutual interaction takes place between the electrical and the thermal field within the semiconductor device. Consequently, by defining margins of the thermally stable/unstable operating regime, the electro-thermal interaction may seriously limit the device lifetime. Therefore, today's semiconductor manufacturers particularly focus their attention on the electro-thermal robustness and reliability of power MOSFETs integrated in modern SP switches in order to guarantee a long operating lifetime during their use in the car environment.

In this chapter basic concepts of modern power MOSFETs integrated in SP switches will be described. Afterwards, their Safe Operating Area (SOA) in forward-biasing operations will be briefly recalled and concepts of reliability and robustness will be introduced accordingly. The second section deals with the behavioral modeling of the power MOSFET, specifically focusing the attention on the temperature dependencies of the main device parameters, which are responsible for the thermal instability effect observed in such transistors. Finally, some basic notions concerning modern SP switches will be given together with their typical critical working conditions commonly experienced during their operation within the automotive environment.

1.1 Power MOSFET basics

1.1.1 VD-MOSFET and Trench MOSFET

Contrary to common MOSFETs employed in analog and digital applications (e.g. signal amplification, digital functions and logic gates, etc.), a power MOSFET is a vertical device, namely the current flows from the top towards the bottom of the silicon die by crossing the whole semiconductor substrate. The reason of such a design lies in the higher blocking voltage capability required by standard power applications, which involves a substantially thick epitaxial layer allowing larger maximum breakdown voltages.

Independent of the technology employed to realize the device, a power MOSFET is physically made of a collection of elementary cells connected in parallel, where every cell generally includes two or more elementary transistors. Modern power MOSFETs can include up to several hundreds of thousands cells depending on their active area size. The dimension of one elementary cell defines the *pitch* of the technology. In general, for a defined active area, the smaller the pitch, the lower the R_{ON} of the device, hence allowing higher current capability.

One of the first power MOSFETs developed was the VD-MOSFET device, the elementary cell of such a device is depicted in Figure 1.1. For the *n*-channel configuration (N-MOS), the device structure can be obtained by growing, using an epitaxial process, a low doped n^- silicon layer upon a generally thicker (hundreds of micrometers) highly doped n^+ silicon substrate. Subsequently, the p-base and the n^+-source region, are realized by opportune implantation and diffusion processes. Both regions are self-aligned thanks to the gate polysilicon electrode which has to be realized before the definition of the source and body regions.

A sufficiently large positive bias applied at the gate electrode ($V_{\mathrm{GS}} > V_{\mathrm{th}}$) allows the formation of a lateral channel in the base region just below the gate oxide between the source and the n^- epitaxial region by attracting the minority carriers (electrons) of the base region. In such a way a current path is created for electrons which can then flow from the source (typically biased at a low potential) to the drain (biased at a high potential) terminal. When electrons reach the end of the channel region, they enter into the so called Junction Field Effect Transistor (JFET) region, then the current spreads towards the n^- epitaxial layer entering the *accumulation* region and finally the *drift* region.

In the OFF-state ($0 \leq V_{\mathrm{GS}} \leq V_{\mathrm{th}}$), a high positive drain voltage V_{D} determine the reverse biasing of the base-drift junction. Since the donor con-

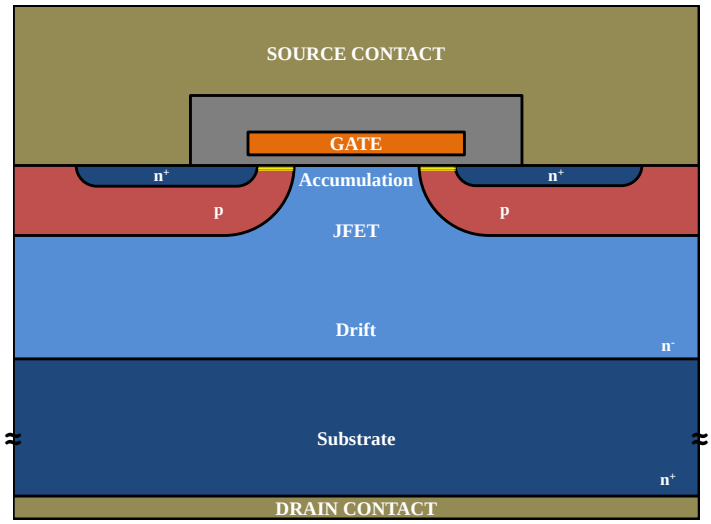

Figure 1.1: VD-MOSFET structure highlighting the critical region for the electron current path.

centration in the drift region is generally substantially lower than the acceptor concentration in the base region, most of the V_{DS} voltage will drop across the drift layer of the device because the space depletion region mainly develops towards the lower doped semiconductor side. Therefore, the role of the drift region is of paramount importance because, by means of opportunely designed doping concentration and thickness, it allows the definition of the maximum voltage V_{DS} which the device can sustain without running into the avalanche effect determined by the impact ionization phenomenon [43].

Unfortunately, a thick and poorly doped drift layer determines also a high internal drift resistance, hence a higher electrical power dissipation which reduces the energy capability of the device. Consequently, when designing the thickness and the doping concentration of the drift region in a VD-MOSFET, the trade-off between its maximum allowed voltage and its R_{ON} must be always taken into account, keeping in mind that in general higher blocking voltage capability implies higher R_{ON} and vice-versa.

An alternative power MOSFET structure which allows to reduce the overall ON-state resistance by substantially shrinking the cell pitch is the Trench power MOSFET, also called *U-MOSFET*. The elementary cell of a trench MOSFET is depicted in Figure 1.2. From the technological viewpoint the trench is realized by etching trench-stripes onto the upper surface of the silicon

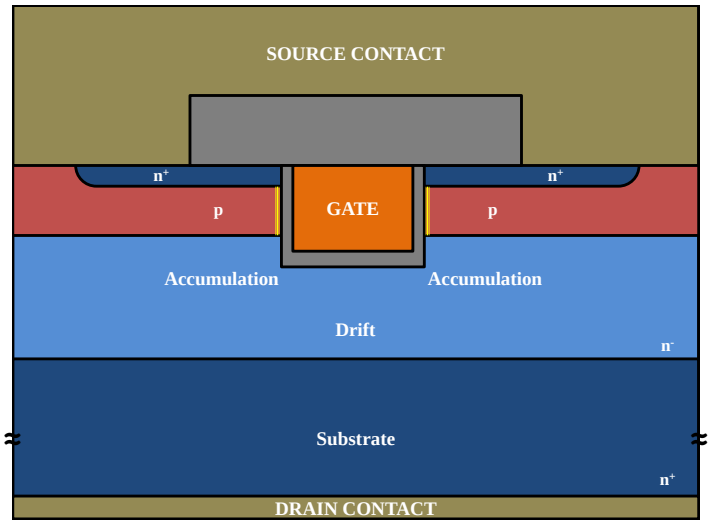

Figure 1.2: Trench MOSFET structure highlighting the critical region for the electron current path. The reader should notice the absence of the JFET region.

die. Subsequently, following opportune deposition processes, those trenches are filled with silicon oxide and polysilicon.

In an *n*-channel power trench MOSFET, by applying an adequate positive gate potential, a vertical inversion channel can be created within the *p*-base region along the vertical trench sidewalls nearby the thin trench oxide. When a positive drain voltage is applied, such a channel provides a path for the electron current to flow from the source towards the drain terminal.

Concerning its blocking voltage capability, in a power trench MOSFET the considerations described above for the VD-MOSFET are valid as well, determining the same trade-off between R_{ON} and breakdown voltage.

Compared with a VD-MOSFET, the real advantage of a trench MOSFET is given by the completely absence of the JFET region which yields a substantial reduction of the overall device R_{ON} without decreasing its blocking voltage capability. For this reason, the trench technology is nowadays widely employed for producing modern discrete power MOSFETs, as well as MOSFETs integrated in current SP switches.

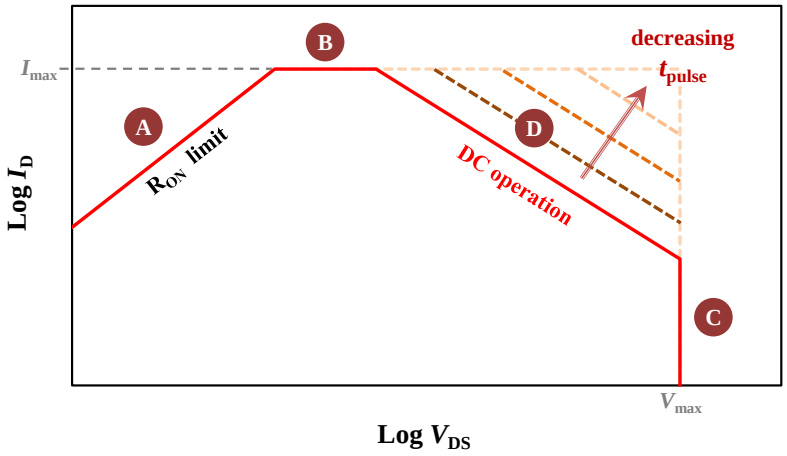

Figure 1.3: SOA of a power MOSFET in DC and pulsed operation.

1.1.2 Safe operating area (SOA)

The SOA of a power device defines the maximum allowed V_{DS} together with the maximum allowed I_{D} at which the device can be operated without destructive failure. In Figure 1.3, the typical SOA of a power MOSFET is depicted using logarithmic scales for both V_{DS} and I_{D}. The limits visible in Figure 1.3 define the border of the biasing area where the device can reliably work without experiencing a failure. In the following, a short description of those limits is given:

- The region A is intrinsically defined by the power MOSFET operation. In this region device operation is limited by the V_{DS} drop, which is defined by its R_{ON} at low-current operations (namely the R_{ON} in the *triode region*);
- The region B is a consequence of the maximum drain current I_{D} allowed for the device. This limit is typically related to the maximum current that device source bond wires can sustain without fusing;
- The region C determines the maximum V_{DS} blocking voltage that the device can handle without experiencing avalanche failures;
- Finally, the region D is mainly determined by the maximum allowed junction temperature $T_{\mathrm{j_{max}}}$, which is defined by device biasing condi-

tions and its junction to case thermal impedance $Z_{\text{th}_{\text{j-c}}}$ by means of:

$$T_{\text{j}} - T_{\text{c}} = Z_{\text{th}_{\text{j}-\text{c}}} V_{\text{DS}} I_{\text{D}} \tag{1.1}$$

where T_{c} is the case (package) temperature which is commonly equal to the ambient temperature. In this region, the range of maximum allowed V_{DS} and I_{D} is eventually larger in pulse operation. In fact, in this case a reduced duty cycle implies a smaller effective power dissipation compared to the one obtained in the DC operation. Therefore, this SOA limit enlarges as the pulse width is reduced.

The definition of the SOA limits depicted in Figure 1.3 is based on just purely thermal considerations and neglects possible further limitations related to electro-thermal interactions which commonly take place within a power MOSFET. We will see in the following of this chapter (specifically in Paragraph 1.2.3) that a further region in the MOSFET SOA emerges due to the thermal instability regime which typically characterizes the device behavior in high-voltage/low-current operations.

The reader can find a highly recommended extensive study about the single-stress SOA of power MOSFETs integrated in SP switches in the PhD thesis of Marie Denison [44].

1.1.3 Robustness and reliability

Power semiconductor products available on the market have intrinsic performance limits and, during their lifetime, they are obviously subjected to wear. Additionally, within a certain specific customer application their operating limits may be very close to application specifications. Based on these considerations, two basics concept have been defined for semiconductor products, i.e. the robustness and the reliability. Their notions are often diverging, therefore in this paragraph they will be briefly defined.

The robustness, also called *ruggedness*, of a power device refers to its capability to handle operation beyond normal conditions and device specifications. Assessing the robustness of a power transistor means to verify that the device is capable to perform its intended functions with sufficient safety margins (i.e. robustness margins) between specification limits and one or more failure mechanisms in a defined application [45, 46]. In Figure 1.4, device specifications of two generic parameters, i.e. A and B, have been schematically represented with an ellipse which must completely contain the application requirements (depicted by means of a rectangle). Beyond its robustness margin the device may fail due to a certain failure mechanism.

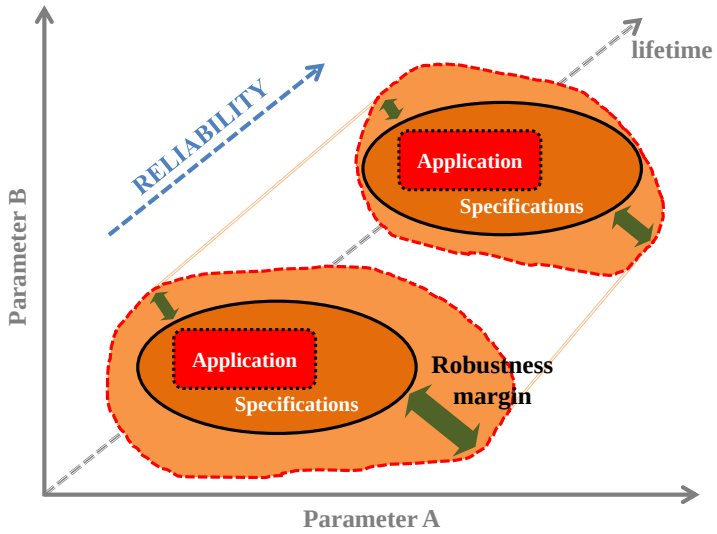

Figure 1.4: Schematic representation of both reliability and robustness concepts [46]. The usage of the device determines the shrinking of its robustness margins and, consequently, a lower reliability during its lifetime.

Reliability, instead, is defined as "the probability of a device to perform a required function under given conditions for the expected lifetime" [6].

The usage of the product in its lifetime within a specified application may eventually affects both its robustness margins and, eventually, its specifications. Contrary to the definition of robustness, the reliability is not an inherent property of the device, but it is rather related to the application requirements. Hence, validating the reliability means verifying that the device is capable to fulfill the application requirements within the specified lifetime.

Generally, both the robustness and the reliability of power devices are experimentally evaluated. In case of reliability estimations, repetitive tests help to predict the lifetime of the device by means of opportune statistical analyses [47, 48]. The reader can find examples of test systems purposely designed for assessing the reliability of products for automotive applications in [49, 50, 51, 52, 53]. Finally, analysis of those parameters returns important information on both failure and degradation mechanisms which helps semiconductor manufacturers to improve their products and to provide accurate product specifications, i.e. datasheets, useful for their customer applications.

In this work an electro-thermal simulator will be presented in detail. This tool is particularly effective for estimating the robustness of power MOSFETs under critical operating conditions, such as Short Circuit (SC) or switching of inductive loads. Nevertheless, in the second part of Chapter 4, a peculiar reliability characteristic related to a specific failure mechanism observed in modern SP devices will also be analysed by means of the developed simulator.

1.2 Power MOSFET modeling

1.2.1 Analytical modeling of power MOSFETs

In this paragraph a few elementary analytical models describing the power MOSFET electrical behavior will be introduced. Contrary to standard small-signal MOSFETs which are four-terminal devices, a power MOSFET features only three terminals because in its structure the source and the body region are always connected together (see the VD-MOSFET and the trench MOSFET structures reported in Figure 1.1 and in Figure 1.2, respectively). Models proposed in the following of this section are valid for every kind of power MOSFET, namely they are independent on the technology used to realize it.

It can be demonstrated that starting from the Poisson's equation describing the potential distribution within the device channel and the Gauss's law for defining the space charge in the semiconductor, the infinitesimal voltage drop $\mathrm{d}V$ within the channel of an N-MOS elementary cell can always be written as a function of the cell drain current I_{D} as follows [43, 54]:

$$\mathrm{d}V = I_{\mathrm{D}}\mathrm{d}R = I_{\mathrm{D}}\frac{\mathrm{d}x}{W_{\mathrm{cell}}\mu_n C_{\mathrm{OX}}\left[V_{\mathrm{GS}} - V_{\mathrm{th}} - V(x)\right]} \tag{1.2}$$

where $\mathrm{d}R$ represents the infinitesimal resistance in the device channel which depends on:

- the channel width W_{cell} of an elementary cell of the device;
- the electron mobility in the channel μ_{n};
- the gate oxide (specific) capacitance, which is a function of the permittivity within the oxide ϵ_{OX} and the oxide thickness t_{OX} by means of:

$$C_{\mathrm{OX}} = \frac{\epsilon_{\mathrm{OX}}}{t_{\mathrm{OX}}} \tag{1.3}$$

- the gate-source voltage drop V_{GS};

- the threshold voltage V_{th}, which can be expressed as:

$$V_{\text{th}} = \frac{\sqrt{4\epsilon_{\text{S}}\text{k}TN_{\text{A}}\ln\left(\frac{N_{\text{A}}}{n_{\text{i}}}\right)}}{C_{\text{OX}}} + 2\frac{\text{k}T}{\text{q}}\ln\left(\frac{N_{\text{A}}}{n_{\text{i}}}\right) \tag{1.4}$$

  where ϵ_{s} is the permittivity of silicon, k the Boltzmann constant, T the absolute temperature, N_{A} the acceptor doping level in the body region, n_{i} the intrinsic carrier concentration in silicon at the temperature T and q is the elementary charge.

- the potential $V(x)$ representing the local potential in the channel along the direction of the current.

By integrating Equation 1.2 along the entire channel length L_{ch}, namely:

$$\int_0^{L_{\text{ch}}} I_{\text{D}}\text{d}x = W_{\text{cell}}\mu_n C_{\text{OX}}\int_0^{V_{\text{DS}}}\left[V_{\text{GS}} - V_{\text{th}} - V(x)\right]\text{d}V \tag{1.5}$$

an expression for the drain current in the channel can be obtained:

$$I_{\text{D}} = \frac{1}{2}\mu_n C_{\text{OX}}\frac{W_{\text{cell}}}{L_{\text{ch}}}\left[2\left(V_{\text{GS}} - V_{\text{th}}\right)V_{\text{DS}} - {V_{\text{DS}}}^2\right] \tag{1.6}$$

For a power MOSFET, generally three different regimes of operations can be defined:

- interdiction ($V_{\text{GS}} < V_{\text{th}}$) , i.e. for low biasing gate potentials the channel cannot be created;

- triode, or linear, region ($V_{\text{GS}} \geq V_{\text{th}}$ and $V_{\text{GS}} - V_{\text{th}} \geq V_{\text{DS}}$), i.e. for biasing gate potentials sufficiently high an inversion layer (channel) is created in the body region;

- saturation, or pinch-off, region ($V_{\text{GS}} \geq V_{\text{th}}$ and $V_{\text{GS}} - V_{\text{th}} < V_{\text{DS}}$), i.e. for even higher biasing gate potentials an inversion layer (channel) is created and the pinch-off effect takes place.

According to operating regimes defined above, the Equation 1.6 can be rewrit-

ten as follows:

$$I_{\mathrm{D}} = \begin{cases} \text{if } \; V_{\mathrm{GS}} < V_{\mathrm{th}} : \\ 0 \\ \\ \text{if } \; V_{\mathrm{GS}} \geq V_{\mathrm{th}} \; \text{ and } \; V_{\mathrm{GS}} - V_{\mathrm{th}} \geq V_{\mathrm{DS}} : \\ \frac{1}{2}\mu_{\mathrm{n}} C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left[2 \left(V_{\mathrm{GS}} - V_{\mathrm{th}} \right) V_{\mathrm{DS}} - V_{\mathrm{DS}}^2 \right] \\ \\ \text{if } \; V_{\mathrm{GS}} \geq V_{\mathrm{th}} \; \text{ and } \; V_{\mathrm{GS}} - V_{\mathrm{th}} < V_{\mathrm{DS}} : \\ \frac{1}{2}\mu_{\mathrm{n}} C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left[V_{\mathrm{GS}} - V_{\mathrm{th}} \right]^2 \end{cases} \tag{1.7}$$

This equation represents the easiest way to analytically describe the electrical device behavior at its terminals. A slightly modified version of Equation 1.7 includes a multiplicative term which takes into account the *channel modulation effect*. Such an equation is often referred as the *SPICE level-1 model* of the MOSFET due to the fact that it is commonly implemented in circuit simulators, like SPICE [55]. In the next paragraph we will see that the electrical behavior described by Equation 1.7 is already sufficiently accurate when simulating the electro-thermal interaction of a power MOSFET operating in a limited V_{GS}-range. Nevertheless, more accurate models can be used for defining the electrical behavior of the elementary cell, such as higher order SPICE models (e.g. SPICE level-3), though they are usually quite complex because they imply the definition of many parameters.

In this work, the attention has been focused on the usage of elementary electrical (and thermal) representations which can describe with a satisfactory level of accuracy the relatively simple cell behavior within a power MOSFET when simulating its electro-thermal response. Specifically, two further analytical models have been used in this research context. In the first one, the SPICE level-1 model has been slightly modified by expressing the source potential V_{S} in the V_{GS} voltage as a function of the distributed source resistance R_{s}. This modeling is particularly suitable for describing the cell behavior in devices with a large active area or with a thick power metallization. In fact, in those MOSFETs the *de-biasing effect* may be relevant, namely the finite value of the electrical conductivity associated to the source metallization may determine a substantially different V_{S} potential at the top and at the bottom of the source power metallization [56, 57, 58]. Consequently, the effective V_{GS} seen by a generic device cell becomes smaller than the one applied to the device electrodes, due to the voltage drop V_{S} across the power metal, which is in turn a

function of the drain current:

$$V_{\mathrm{GS}} = V_{\mathrm{G}} - V_{\mathrm{S}} = V_{\mathrm{G}} - R_{\mathrm{S}} I_{\mathrm{D}} \tag{1.8}$$

Therefore, by substituting Equation 1.8 in Equation 1.7, the cell current in saturation region becomes:

$$I_{\mathrm{D}} = \frac{1}{2} \mu_{\mathrm{n}} C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left[V_{\mathrm{G}} - R_{\mathrm{s}} I_{\mathrm{D}} - V_{\mathrm{th}} \right]^2 \tag{1.9}$$

where R_{s} becomes a parameter for the modified SPICE level-1 model.

Another model which achieves a quite accurate description of the cell I_{D} with a limited number of parameters has been proposed by Enz, Krummenacher and Vittoz [59] and it has been already used by Pfost et al. in [60] for defining the drain current in a VD-MOSFET structure:

$$I_{\mathrm{D}} = I_0 (ABV_{\mathrm{T}})^A \ln \left(1 + \mathrm{e}^{\frac{V_{\mathrm{GS}} - V_{\mathrm{th}}}{ABV_{\mathrm{T}}}} \right)^A \tag{1.10}$$

where I_0 is a constant, V_{T} the thermal voltage, A and B are set to 2.0 and 1.0 respectively, but they can be both used as parameters for achieving higher accuracy when modeling the typically large drift region of a power MOSFET.

Finally, an accurate, though more complex, model can be found in the PhD thesis of d'Alessandro [61], where the electrical (and thermal) response of the elementary cell of a VD-MOSFET has been described using an extended version of the SPICE level-3 model. Such a modeling takes into account the *quasi-saturation effect* which affects a power MOSFET operating in the triode region at high current levels [62, 63]. Due to the velocity saturation phenomena observed in carriers crossing the epitaxial layer at high electric fields, the drift resistance R_{drift} cannot be considered independent of the voltage drop across the epitaxial layer, hence it is modeled as follow:

$$R_{\mathrm{drift}}(V_{\mathrm{DS}}) = R_0 + R_1 \frac{V_{\mathrm{DS}}}{V_{\mathrm{C}} + V_{\mathrm{DS}}} \tag{1.11}$$

where R_0, R_1 and V_{C} represent three parameters. The contribution of R_{drift} can be simply added to the channel resistance within the equation of the drain current because those two resistances are connected in series.

1.2.2 Main temperature dependencies

The analytical models presented in the previous paragraph are suitable for describing the electrical behavior of the elementary cell within a power MOSFET. The electro-thermal interaction taking place in such devices is a direct

consequence of the temperature dependence of the drain current resulting from the electron mobility and the threshold voltage.

The electron mobility within the channel is a decreasing function of the temperature because at high temperature the lattice scattering prevails on the impurity scattering. Its temperature dependence can be described by means of a power law model [43]:

$$\mu_{\mathrm{n}}(T) = \mu_0 \left(\frac{T}{T_0}\right)^{-M} \tag{1.12}$$

where T_0 is a reference temperature, e.g. the ambient temperature, and μ_0 is the value of the mobility at the reference temperature. Both M and μ_0 represent two parameters for the model.

On the other hand, due to the dependencies of the threshold voltage on the Fermi potential and on the work-function, also V_{th} is a decreasing function of the temperature. A commonly adopted model prescribes a linear decrease of V_{th} for increasing temperature [64]:

$$V_{\mathrm{th}}(T) = V_{\mathrm{th_0}} - \Phi\,(T - T_0) \tag{1.13}$$

where $V_{\mathrm{th_0}}$ is the value of the threshold voltage at the reference temperature T_0 and Φ is its temperature coefficient (defined positive in Equation 1.13 due to the minus sign).

In the case of the simple SPICE level-1 model, the temperature dependence of the cell current can be obtained by substituting both Equation 1.12 and Equation 1.13 in Equation 1.7. For example, the drain current in the saturation region can be rewritten in the following way:

$$I_{\mathrm{D}}(T) = \frac{1}{2}\left[\mu_0 \left(\frac{T}{T_0}\right)^{-M}\right] C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left\{V_{\mathrm{GS}} - \left[V_{\mathrm{th_0}} - \Phi\,(T - T_0)\right]\right\}^2 \tag{1.14}$$

The consequences of such temperature dependencies on the device thermal behavior will be described in the next paragraph.

Tabulated values of I_{D} for different temperatures and different biasing conditions have been obtained by means of so called *TCAD simulations* (or *device simulations*). With this kind of simulations, equations describing the drift-diffusion of carriers in doped semiconductors are solved and computed using a numerical method (typically Finite Element Method (FEM)). The commercial tool *Taurus MEDICI* together with the associated process simulator *Taurus T-SUPREM4* allowed to estimate current values in the elementary cell of the

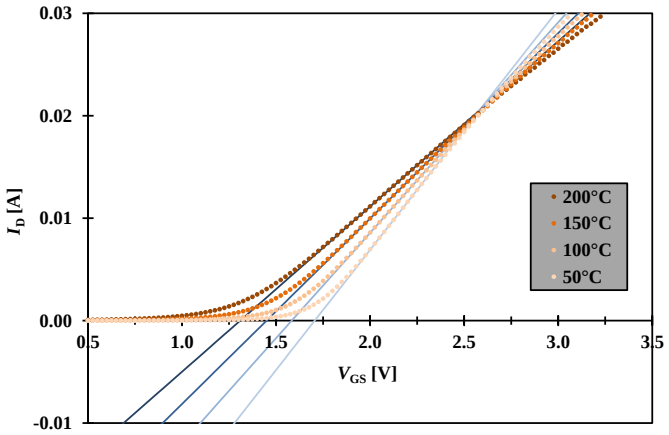

Figure 1.5: Example of threshold voltage extrapolation at different ambient temperatures using the quadratic extrapolation method. The intersection of the obtained line with the V_{GS}-axis returns the value of V_{th} for the considered ambient temperature.

considered technology for a wide range of temperatures, V_{GS} and V_{DS} voltages. Those tabulated values will be referenced in the course of this thesis as *table model* of the device elementary cell. Subsequently, analytical models presented in the previous paragraph have been calibrated by fitting the table model obtained from TCAD simulations.

For the SPICE level-1 model (Equation 1.7) and its modified version (Equation 1.9), the temperature dependence of the threshold voltage has been included using the linear function defined by Equation 1.13. For this purpose, the *quadratic extrapolation method* [65] has been used in order to extract, from the table model, values of V_{th} at different temperatures (see Figure 1.5). Subsequently, the extracted threshold voltages have been fitted using the model of Equation 1.13 and values for V_{th_0} and Φ have been deduced from the intercept and slope of the linear function . Extracted values of the threshold voltage at different temperatures and their resulting linear fitting have been both plotted in Figure 1.6.

The second step consisted in the fitting of tabled temperature dependent transfer characteristics by means of the equation of the drain current in which the temperature dependence for V_{th} has been included, i.e.:

$$I_{\mathrm{D}}(T) = \frac{1}{2}\mu_{\mathrm{n}}(T)C_{\mathrm{OX}}\frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}}\left\{V_{\mathrm{GS}} - \left[\tilde{V_{\mathrm{th}_0}} - \tilde{\Phi}\,(T - T_0)\right]\right\}^2 \qquad (1.15)$$

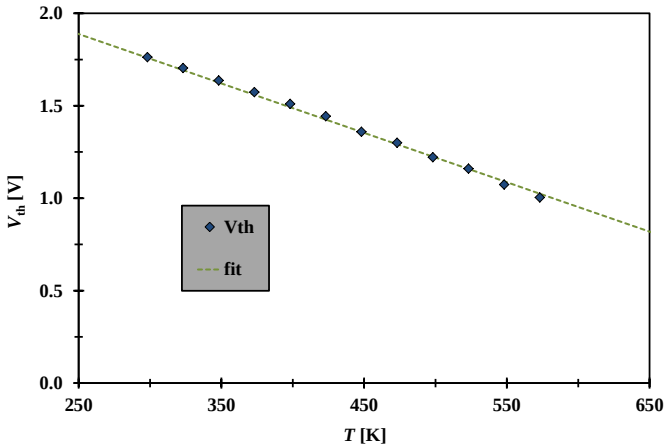

Figure 1.6: Interpolation by linear model (Equation 1.13) of extrapolated threshold voltages at different temperatures.

where $\tilde{\Phi}$ and $\tilde{V_{\mathrm{th}_0}}$ have been previously extracted. Here μ_{n} has been considered as a fitting parameter for every temperature value. For both the SPICE level-1 models presented, μ_{n} values have been extracted from temperature dependent transfer characteristics in a limited V_{GS}-range. In particular, in the case of its modified version (i.e. Equation 1.9), also R_{S} has been used as a fitting parameter. Finally, mobilities obtained at different temperatures have been fitted using the model described in Equation 1.12, determining values for μ and M.

For instance, in Figure 1.7 table model transfer characteristics for V_{DS} values of 10 V, 20 V, 30 V and 40 V at 100°C are compared with the three mentioned analytical models: i.e. SPICE level-1, SPICE level-1 with R_{S} and EKV model. For all of them, the respective parameters have been obtained by fitting the table model data within a V_{GS} range from V_{th} to 3.5 V.

In Chapter 2 a FEM-based electro-thermal simulator will be introduced. With such a simulator, the device transient electrical and thermal responses can be simulated under arbitrary biasing conditions. Several simulations have verified that simple electro-thermal analytical models presented above allow a quite accurate description of the cell current in a limited range of V_{GS} (few volts above the V_{th}). For example, in Figure 1.8 the plotted simulation results refer to the same power MOSFET and operating conditions and have been obtained using for the cell current both the table model description and the simple SPICE level-1 model. Both the simulated current and temperature hot-

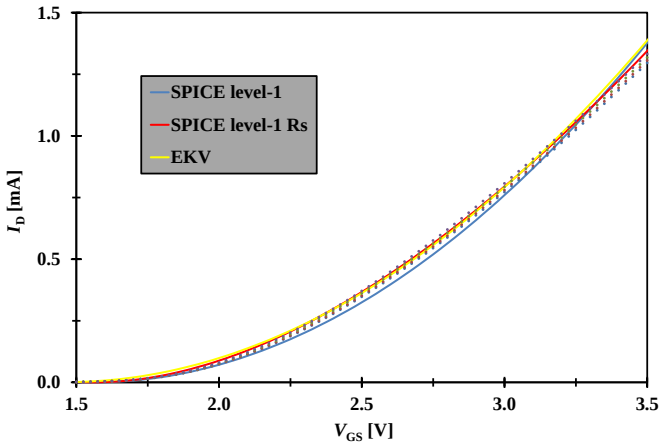

Figure 1.7: Table model transfer characteristics at 100°C for V_{DS} values of 10 V, 20 V, 30 V and 40 V (dots) compared with three different analytical models: SPICE level-1 (blue curve), SPICE level-1 with R_S (red curve) and EKV model (yellow curve) resulting from fitting in a V_{GS} range from V_{th} to 3.5 V.

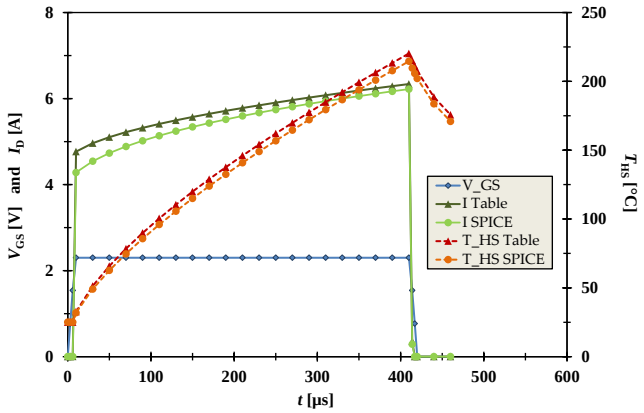

Figure 1.8: Current (green curves) and hot-spot temperature (red/orange dashed curves) obtained from electro-thermal simulations of a power MOSFET. The same operating condition have been simulated using two different models for describing the cell current: table model (triangles) and SPICE level-1 model (dots).

spot evolution obtained with the two cell current descriptions show relatively small deviations. In particular, the maximum I_{D} deviation is 0.5 A and appears at the beginning of the SC, while the maximum T_{HS} deviation of 5.7 K is observed at the end of the pulse.

1.2.3 Thermal instability

The reason for the thermal instability phenomenon in power MOSFETs can be discovered by having a careful look at the SPICE level-1 equation for the saturation regime, which has been reported below by highlighting temperature dependencies of both the μ_{n} and the V_{th}:

$$I_{\mathrm{D}}(T) = \frac{1}{2}\mu_{\mathrm{n}}(T)C_{\mathrm{OX}}\frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}}\left[V_{\mathrm{GS}} - V_{\mathrm{th}}(T)\right]^2 \tag{1.16}$$

For a defined operating condition, i.e. for a fixed V_{GS}, due to Equation 1.12 and Equation 1.13 both the channel mobility and the threshold voltage decrease for increasing temperatures determining two counteracting effects on the drain current. If the effect on I_{D} of the mobility decreasing at increasing temperatures dominates the decrease of the threshold voltage, the overall drain current in Equation 1.16 diminishes. As a consequence, the so called *temperature coefficient of the current* α is negative, where α is given by:

$$\alpha = \frac{\mathrm{d}I_{\mathrm{D}}}{\mathrm{d}T} \tag{1.17}$$

On the contrary, if for increasing T the decrease of the threshold voltage dominates the decrease of the mobility, then I_{D} increases determining in this case a positive α. Those two counteracting mechanisms define two different V_{GS}-ranges which can be noticed in temperature dependent transfer characteristics reported in Figure 1.9.

In general, for low V_{GS} voltages, i.e. at low current levels, the latter mechanism is dominant, while at high V_{GS} voltages, i.e. at higher current levels, the former mechanism prevails on the latter one. The border between those two ranges is given by the V_{GS} voltage value for which the decrease of the mobility with the temperature is completely counteracted by the decrease of the threshold voltage, determining a value of the drain current which is substantially independent of temperature. For this reason, the pair given by V_{TCP} and I_{TCP} in Figure 1.9 defines the so called TCP. Due to the fact that for this point α is equal to zero, the TCP is often referred also with the name of Zero Temperature Coefficient (ZTC) point.

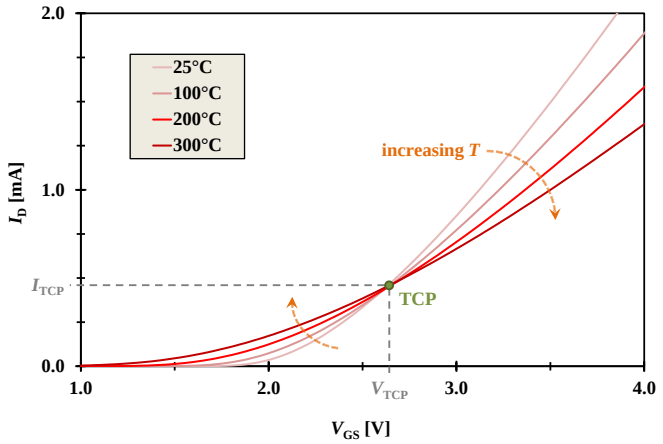

Figure 1.9: Cell current at different ambient temperatures and definition of the TCP. Two distinct operating regimes can be highlighted: a thermally unstable one for V_{GS}<V_{TCP} (or I_D<I_{TCP}) and a thermally stable one for V_{GS}>V_{TCP} (or I_D>I_{TCP}).

As aforementioned, for V_{GS} lower than V_{TCP} (and larger than V_{th}), α is positive. This means that a local increase of temperature determines also a local increase of current. Consequently, the increase of current leads to a local increase of power within the cell which induces a further temperature rise resulting, in turn, in a positive feedback between the current and the temperature increase. If such a mechanism takes place within the device for a sufficient time, extremely high temperatures may be reached within the silicon and they may lead the device to thermal runaway. Therefore, this kind of operation is typically referred as the *thermally unstable regime* and the related effect is commonly called thermal instability. In this regime, an opportune control on device biasing conditions is extremely important in order to avoid possible failures. On the other hand, for V_{GS} larger than V_{TCP}, the drain current decreases at increasing temperature due to its negative temperature coefficient. As a consequence, an eventual increase of temperature within the elementary cell is balanced by a decrease of current and power which, in turn, tend to decrease the temperature. Therefore, this kind of operation is defined as the *thermally stable regime*. In standard applications, power MOSFET biasing conditions are typically chosen within their thermally stable operating range. Nevertheless, even for such conditions a power MOSFET may experience the thermal runaway effect. In fact, for too high current levels (imposed by the electric

load) the power dissipation within silicon may be large enough to determine a severe temperature increase within the device. Consequent temperature peaks can be high enough so that even the counteracting effect of the thermal stability would not be capable to limit them. Anyhow, in those biasing situations, the thermal runaway phenomenon is defined by only thermal considerations and may determine device failures which are commonly taken into account within the SOA.

A more general criterion for defining the thermal stability in a power MOSFET takes into account also the V_{DS} voltage and it has been proposed [12, 66] and experimentally validated [67] by Spirito et al.. Based on [68], the thermal instability condition for a power device is usually defined by:

$$\frac{\partial P_{\mathrm{gen}}}{\partial T} \geq \frac{\partial P_{\mathrm{diss}}}{\partial T} \tag{1.18}$$

where P_{gen} and P_{diss} represent the electrically generated and thermally dissipated power, respectively. The first term of Equation 1.18 can be rewritten by recalling the definition of the power generated in a device due to the Joule effect and using the definition of α:

$$\frac{\partial P_{\mathrm{gen}}}{\partial T} = V_{\mathrm{DS}} \frac{\partial I_{\mathrm{D}}}{\partial T} = V_{\mathrm{DS}}\, \alpha \tag{1.19}$$

Furthermore, the thermally dissipated power term P_{diss} is defined by the thermal impedance $Z_{\mathrm{th_{j\text{-}c}}}$ and the temperature difference between the junction temperature T_{j} and the case temperature T_{c}:

$$P_{\mathrm{diss}} = Z_{\mathrm{th_{j-c}}} \Delta T_{\mathrm{j-c}} \tag{1.20}$$

Therefore, Equation 1.18 can be rewritten by substituting Equation 1.19 and Equation 1.20, i.e.:

$$\alpha \geq \frac{1}{V_{\mathrm{DS}} R_{\mathrm{th_{j-c}}}} \tag{1.21}$$

The equation above represents a very general criterion describing the onset of the thermal instability in a power MOSFET based on its biasing conditions and on its thermal impedance.

It has been experimentally observed that in power MOSFETs biased under thermally unstable conditions two effects take place at the same time, namely the progressive formation of a temperature hot-spot and the tendency of the current density to focus towards the hot-spot location (usually called

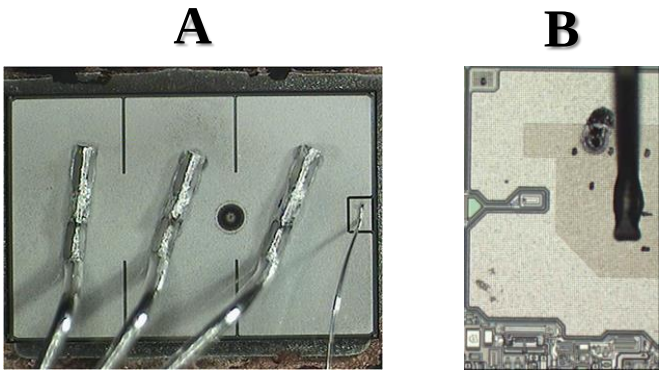

Figure 1.10: Burn mark caused by a thermally unstable biasing condition in a discrete power MOSFET (A) and in a SP switch (B).

current focusing, or *current crowding*, effect) [12, 31, 69, 20]. Those two effects "feed" each other due to the electro-thermal interaction which takes place within device elementary cells. If the unstable biasing condition is applied to the device for a sufficient time, a failure described by a localized burn mark is typically observed as in Figure 1.10. The physics-based mechanism leading to such a failure, which will be described in detail in the next chapter (in Paragraph 2.1.1), determines a further reduction of the power MOSFET SOA [12, 13], as depicted in Figure 1.11. It can be noticed that the SOA limitation takes place at high voltages and low current levels resembling the *second breakdown* phenomenon observed in power BJTs, although here the reasons causing such effect are different.

In conclusion, we have seen that this failure mechanism is determined by the electro-thermal interaction which takes place within the elementary cells of a power MOSFET. Consequently, its correct modeling and simulation becomes crucial in order to accurately predict the robustness characteristics of power MOSFETs.

1.2.4 Modeling the parasitic bipolar transistor

By having a close look on the structure of the elementary cell of both a power VD-MOSFET or a trench MOSFET, a bipolar *n-p-n* transistor defined by the drain-body-source structure can be discovered. For example, the parasitic BJT has been highlighted in the elementary cell of a trench MOSFET in Fig-

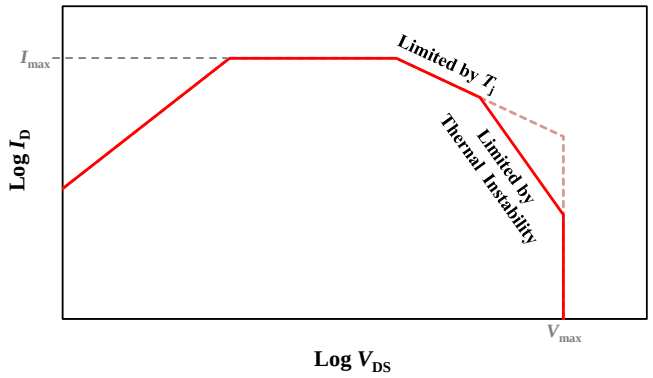

Figure 1.11: Effect of the thermal instability on the SOA of a power MOSFET. At high-voltage/low-current levels, the region limited by the maximum allowed junction temperature is further shrunk due to the thermal instability phenomenon.

ure 1.12-A. From a circuit point of view, such a bipolar transistor is connected to the main Diffused MOSFET (DMOS) in a parallel manner as depicted in Figure 1.12-B, i.e.:

$$I_{\mathrm{D}} = I_{\mathrm{MOS}} + I_{\mathrm{BJT}} \tag{1.22}$$

where I_{MOS} is the drain current of the elementary cell already described in Paragraph 1.2.1, while I_{BJT} is the current associated to the parasitic bipolar transistor.

The role of the BJT is particularly relevant in a power MOSFET because its eventual activation may lead the whole power device to failure due to the *latch-up effect*. In fact, if the current flowing within the parasitic base resistance R_{B} determines a large enough V_{BE} voltage drop across the base-emitter junction of the BJT, a substantially higher and uncontrolled current would flow across the device cell increasing its local junction temperature T_{j}. Furthermore, the temperature increase determines an increasing current gain of the BJT (the β parameter has a positive temperature coefficient) which would further increase the parasitic component of the MOSFET current (I_{BJT}), resulting in turn in a positive, and destructive, feedback between I_{BJT} and T_{j}.

In this work, both simulations and experimental measurements have been performed using a trench technology developed by Infineon Technologies [70]

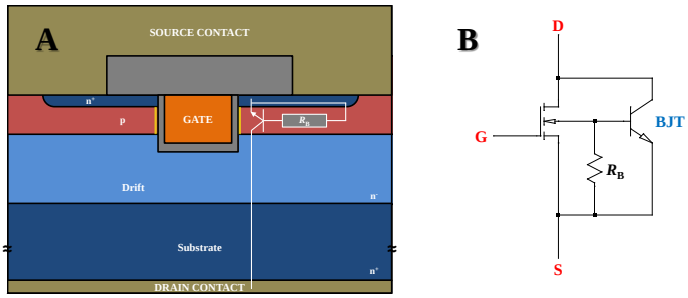

Figure 1.12: Parasitic bipolar transistor (*n*-*p*-*n*) within the elementary cell of a power trench MOSFET (A) and equivalent circuit including the base resistance R_{B} (B).

in which, due to a very low internal base resistance and low impact ionization, the parasitic BJT never enters the active operation regime during relevant pulses. Nevertheless, even in such a robust design, the parasitic structure may accelerate the device failure mechanism at high temperature, due to its leakage current contribution which: 1) is dependent on the junction temperature with a positive temperature coefficient; 2) becomes relevant at very high temperature (i.e. for $T_{\mathrm{j}} > 350°\mathrm{C} \sim 400°\mathrm{C}$).

The current density J associated to the parasitic bipolar structure is given by the sum of several contributions:

$$J_{\mathrm{BJT}} = J_{\mathrm{av}} + J_{n\mathrm{E}} + J_{\mathrm{CB}} \simeq J_{n\mathrm{E}} + J_{\mathrm{CB}} \tag{1.23}$$

In Equation 1.23 the J_{av} term describes the current contribution related to the electron-hole pairs generated in the body-drain depletion region by impact ionization. In this technology, it has been demonstrated by means of TCAD simulations that, when operating the device at V_{DS} not larger than 40 V, namely when V_{DS} is sufficiently far from the maximum breakdown voltage, this contribution is small enough even at a V_{GS} value of 4 V [25]. Furthermore J_{av} exhibits a negative temperature coefficient, hence its contribution in the Equation 1.23 has been neglected.

The $J_{n\mathrm{E}}$ term describes the current contribution due to the injection of electrons from the emitter into the base-collector junction. According to the Gummel-Poon model, $J_{n\mathrm{E}}$ represents a diffusion component and can be expressed as [43]:

$$J_{n\mathrm{E}}(T) = \frac{\mathrm{q}D_{\mathrm{n}}n_{\mathrm{i}}^2}{\int_0^{L_{\mathrm{B}}} p_{\mathrm{p}}\mathrm{d}x} e^{\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}}} \simeq \frac{\mathrm{q}D_{\mathrm{n}}n_{\mathrm{i}}^2}{\int_0^{L_{\mathrm{B}}} p_{\mathrm{p}}\mathrm{d}x} \tag{1.24}$$

where D_{n} represents the electron diffusivity, while the integral term at the denominator is the Gummel number which gives the total impurity dose per area in the base region. In Equation 1.24, the exponential term has been approximated to 1 because, as already aforementioned, in the treated technology design V_{BE} can be neglected due to the very low parasitic base resistance achieved. However, in cases where V_{BE} is not negligible (i.e. the parasitic BJT may become active), a valid model has been reported in [44].

The J_{CB} term models the current contribution associated to the collector-base junction which is reverse biased. In the space charge region of this junction the generation of electron-hole pairs takes place resulting in a majority carrier current component. At the same time, in the neutral region of the collector-base junction, both holes and electrons diffuse resulting in a second current contribution for J_{CB}. Therefore, for J_{CB} term can be expressed in the following way:

$$J_{\mathrm{CB}} = \frac{\mathrm{q}W_{\mathrm{D}}n_{\mathrm{i}}}{\tau_{\mathrm{g}}} + \mathrm{q}n_{\mathrm{i}}^2 f(N_{\mathrm{D}}, N_{\mathrm{A}}) \tag{1.25}$$

where W_{D} and τ_{g} in the first term of Equation 1.25 (drift term) are the depletion layer width and the carrier generation lifetime in the depletion region of the collector-base junction, respectively. The f function in the second term (diffusion term) describes a complex function defining the diffusion component of the electrons and holes based on, acceptor N_{A} and donor N_{D} concentrations, on the carrier lifetime of holes τ_{p} and electrons τ_{n} and on the diffusivity of holes D_{p} and electrons D_{n}.

Both Equation 1.24 and Equation 1.25 exhibit a significant temperature dependence because, n_{i}, D_{p}, D_{n}, τ_{g}, τ_{p} and τ_{p} depend on the temperature. In particular the intrinsic carrier concentration is a well-known function of the temperature:

$$n_{\mathrm{i}}(T) = \sqrt{N_{\mathrm{C}}N_{\mathrm{V}}}\, e^{-\frac{E_{\mathrm{g}}(T)}{2\mathrm{k}T}} \tag{1.26}$$

where N_{C} and N_{V} are the effective density of states in the conduction and valence band, respectively, while E_{g} is the band-gap energy in silicon, which also depends on the temperature by means of:

$$E_{\mathrm{g}}(T) = 1.169 - \frac{4.9 \cdot 10^{-4}\ T^2}{T + 655} \tag{1.27}$$

It can be demonstrated that the temperature dependence of J_{BJT} is mainly given by temperature dependencies established by: 1) $n_{\mathrm{i}}^2(T)$ for both $J_{n\mathrm{E}}(T)$ and the second component of J_{CB}; 2) $n_{\mathrm{i}}(T)$ for the first component of J_{CB}. By

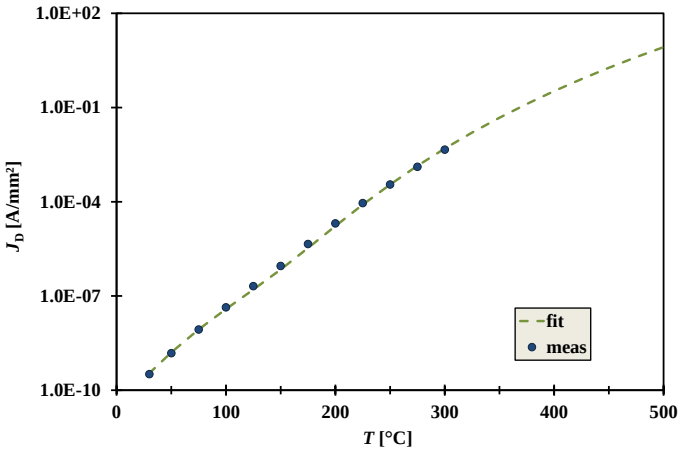

Figure 1.13: Measurements of the leakage current of a technology wafer-level specimen at different ambient temperatures (dots) and fitting with the analytical model of Equation 1.28 (dashed green curve).

explicitating in Equation 1.24 and Equation 1.25 the T variable using Equation 1.26 and Equation 1.27, the temperature dependence of the overall J_{BJT} current density can be expressed as follows:

$$J_{\mathrm{BJT}}(T) = \hat{J}_1(T)\, T^3 e^{-\frac{E_g}{kT}} + \hat{J}_2(T)\, T^{\frac{3}{2}} e^{-\frac{E_g}{2kT}} \tag{1.28}$$

The equation above describes the current density associated to the parasitic bipolar structure, also called leakage current, as a function of the absolute temperature.

Experimental measurements have been carried out in order to fit the J_{BJT} model. The leakage current density of a technology wafer-level specimen have been measured at different ambient temperatures from 25°C to 300°C using a probe station. Experimental values and the fitting curve have been reported in Figure 1.13, where the extrapolation range has been extended up to 500°C because the leakage current associated to a power MOSFET may become significant only at very high temperature. With the available equipment, measurements were possible only up to 300°C, therefore the accuracy of the extrapolated current values has been verified with device simulations. First, the DMOS elementary cell behavior at high temperature has been assessed by means of one-dimensional device simulations (performed using the freeware tool *PC1D* [71, 72]) and afterwards by means of more accurate two-dimensional process and device simulations using *Taurus T-SUPREM4* and

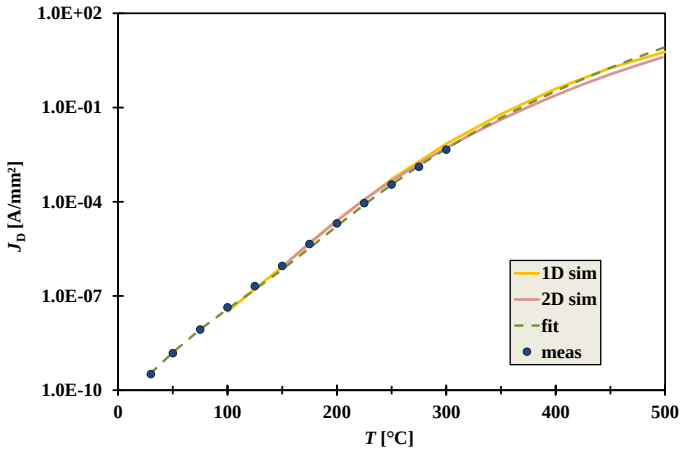

Figure 1.14: Leakage current comparison: measurements (dots), calibrated analytical model (green dashed curve), 1-D device simulation (yellow curve) and 2-D device simulation (orange curve). High temperature current values extrapolated by Equation 1.28 show a good agreement with current values estimated by both device simulations.

Taurus MEDICI. In the Figure 1.14, the obtained simulation results have been plotted together with the experimental data and the fitting curve demonstrating that Equation 1.28 well describes the temperature dependence of J_{BJT} up to 500°C.

The effect of the leakage current in a power MOSFET is of paramount importance for a correct understanding of the device electro-thermal behavior in extreme operations. Due to its exponential temperature dependence, the leakage current contribution becomes relevant only at very high temperature and may lead the device to the thermal runaway failure in its hottest region because of the positive feedback established between current and temperature increase. Therefore, the J_{BJT} contribution is completely uncorrelated with the thermal instability of the drain current described in the previous paragraph, namely a device operating at high T_{j} may fail even for thermally stable biasing conditions. Instead, when operating in the thermally unstable regime, the net effect of the leakage current is to lead the device to an early failure, as it has been demonstrated in Figure 1.15. In this picture, results of an electro-thermal simulation of a 1 mm^2 power MOSFET have been depicted. An unstable biasing operation for the device has been simulated by setting a constant V_{GS} value smaller than V_{TCP} and a high V_{DS}. Two different analytical models have been

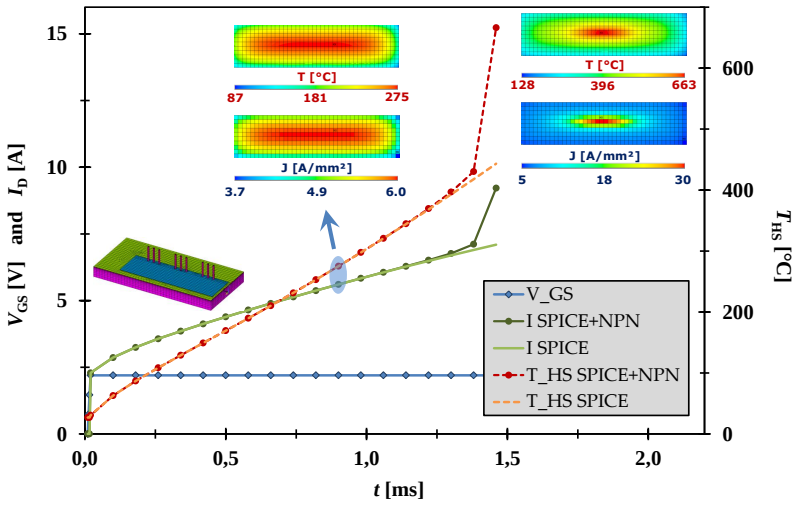

Figure 1.15: Electro-thermal simulations of a power MOSFET performed with two different analytical models of the cell current: SPICE level-1 with and without the leakage current contribution I_{BJT}. The simulated thermal runaway can not be correctly estimated if the cell description would not take into account the effect of the leakage current associated to the parasitic BJT. During the pulse, a hot-spot emerges and, at the same time, the current crowds towards the hot-spot location.

used to define the temperature dependent cell current. The first one features only the SPICE level-1 model including the temperature dependencies of both the V_{th} and the mobility μ_{n} (i.e. Equation 1.7). The second model instead has been obtained by including into the first model the temperature dependent contribution of the leakage current, according to Equation 1.22. The faster rise of both the drain current and the temperature hot-spot obtained by means of the model including the leakage current demonstrates that an accurate and reliable prediction of the device behavior at high temperature must absolutely take into account the effect of the parasitic bipolar even in cases where the BJT does not activate.

In conclusion, the leakage current in a power trench MOSFET can be very relevant especially in large area power devices because its contribution scales with the size of the active area. Moreover, I_{BJT} may represent a crucial failure mechanism for devices operating at high temperature. Therefore, during the

development of a new technology, a careful design of crucial parameters, such as the hole concentration in the base (which defines the Gummel number), becomes fundamental to achieving robust devices.

1.3 Smart Power switches

1.3.1 Smart Power basics

Nowadays, SP devices are the most employed power switches in the car electronic environment. In the last decades they have progressively replaced relays, fuses and discrete circuits used in automotive electronics thanks to their smaller dimensions and reduced costs. Their success is mainly due to: 1) the evolution of more demanding and specific application requirements from cars manufacturers; 2) the fast development of additional car features offering improved comfort and safety functions. Those devices are typically used in simple load switching applications for the activation of resistive (e.g. windows and seat heating), capacitive (e.g. lighting) or inductive (e.g. ABS valves, engine cooling fan) loads or in bridge/half-bridge circuit configurations (e.g. windows lifting systems), further typical automotive applications have been reported in Figure 1.16. A SP device is a power Integrated Circuit (IC) in which one or more power MOSFETs are integrated on the same silicon substrate together with analog and digital circuits suitable for driving and protecting the power devices. The schematic in Figure 1.17 shows a very simplified implementation of the most common protection functions for a high side power switch. Those protection concepts will be briefly described here.

- **Over-current protection**

  Part of the logic circuit of a SP device is devoted to conveniently driving the device when the drain current reaches dangerous levels. Such feature exploits a current sensor integrated within the power MOSFET active area. From a circuit point of view, the current sensor can be considered as a smaller MOSFET connected in parallel to the substantially larger power DMOS and it is realized by using few elementary cells of the whole MOSFET structure. A sense current can be continuously monitored by measuring the voltage drop across a shunt resistor. In this way, the current flowing through the whole power MOSFET can be estimated by means of:

$$I_{\mathrm{D}} = k_{\mathrm{sens}} \frac{V_{\mathrm{shunt}}}{R_{\mathrm{shunt}}} \tag{1.29}$$

Figure 1.16: Common automotive applications for a SP switch.

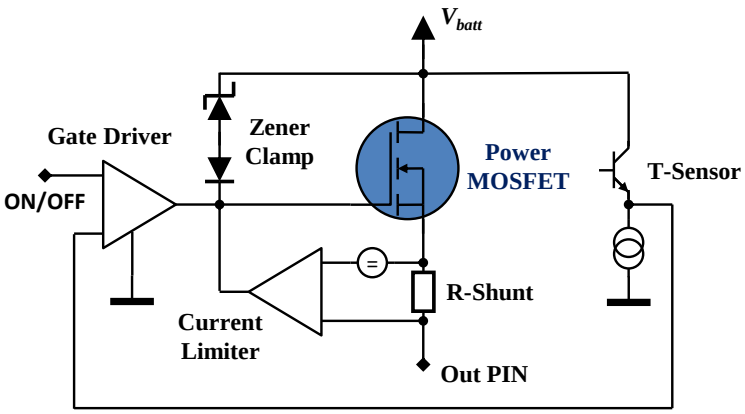

Figure 1.17: Simplified circuit implementation of most common protection functions implemented in a high side SP switch: over-current, over-temperature and over-voltage protection.

where k_{sens} is the ratio between the total number of cells in the power MOSFET and in the sensor MOSFET.

For current levels larger than the maximum allowed current, the gate drive circuitry reduces the V_{GS} of the device, limiting its I_{D} to a safety predefined level. This kind of operation is called *current limitation*.

- **Over-temperature protection**

  In the current limitation operation, the temperature within the device increases even if the power dissipation is limited by the over-current protection mechanism. Consequently, increasing the device temperature may lead the power MOSFET to the failure if not properly controlled. An over-temperature shut-down protection mechanism integrated in the SP device provides to switch-off the MOSFET when a predefined temperature limit is reached. To do so, such strategy needs to continuously measure the DMOS temperature by means of an integrated temperature sensor. Eventually, two temperature sensors may be conveniently used to sense the temperature difference between the ambient and the DMOS temperature, providing a more flexible switch-off strategy which can be effectively used also at different ambient temperatures, e.g. from -40°C to 150°C [73].

  The temperature sensor integrated in the DMOS can be realized following several approaches. In particular, two concepts based on a bipolar structure will be presented in Paragraph 2.4.3, while a very robust bipolar approach and a resistive-based sensor have been reported in [14].

  After switch-off, the device cools down and it is restarted following a hysteresis strategy, namely the DMOS is switched-on again once its sensed temperature reaches an inferior threshold temperature level. During a SC event, after switch-on the temperature will start to increase again and eventually will reach the superior threshold temperature limit which determines its switch-off through the over-temperature protection mechanism. Therefore, during the whole SC pulse the temperature of the device toggles between the two temperature limits as described in Figure 1.18.

- **Over-voltage protection**

  This kind of protection concept is particularly useful for safely switching inductive loads which are responsible of over-voltage spikes during the discharge of their stored magnetic energy. This operation represents

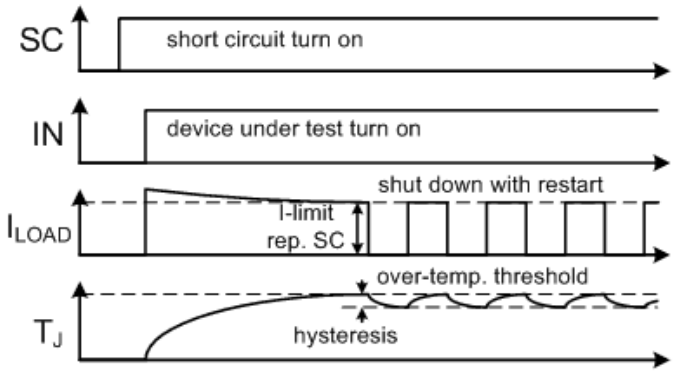

Figure 1.18: Over-temperature protection concept based on the hysteresis strategy: thermal toggling [74].

a critical condition for voltage-unprotected MOSFETs and will be described in detail in the next paragraph. Over-voltage protection concepts are commonly implemented by means of circuitry which avoids avalanche on-set by limiting the source (in high side configuration) or the drain (in low side configuration) potentials. Typical adopted strategies are based on the usage of a clamping network or a freewheeling diode.

- **ESD protection**

  In the harsh electronic automotive environment electrostatic discharges are very frequent, furthermore a SP device may be subjected to Electrostatic Discharge (ESD) pulses even during handling and assembling. These stress conditions may be described by different kind of pulses depending on the source of discharge, such as manufacturing and/or handling machines or human bodies.

  Generally, an ESD event consists of a short high-power pulse which may irreparably damage the DMOS or the logic ICs of the power switch through its pins, hence their limitation is crucial for guaranteeing a reasonable lifetime. This kind of stress pulse can be typically limited by an opportune Z-diode connected between the device pin and the ground. As a consequence, this component must be dimensioned depending on the maximum allowed over-voltage spike that can be sustained on a certain device pin without causing any failure in the SP switch. Other circuit

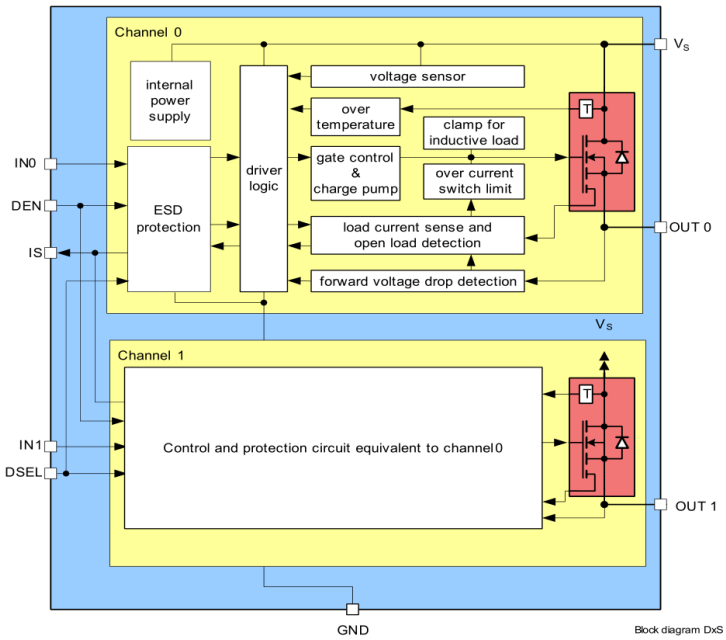

Figure 1.19: Block diagram of the SP device *BTS5020-2EKA* [75].

concepts based on the usage of Z-diodes in combination with smaller power DMOS are also effectively used in modern SP switches.

In addition to the listed protection concepts, SP switches are also equipped with diagnostic functions, such as open-circuit detection, and logic circuits for digital data exchange (through TTL or SPI interfaces) with the car control unit. An example of a commercial SP device currently available on the market is the PROFET *BTS5020-2EKA*. This product features two distinct power MOSFETs (channels), whereas for every device the functions illustrated in the block diagram of Figure 1.19 are provided [75].

Protection functions are necessary to drive (control) the device in a safe manner during critical operating conditions in the harsh automotive environment. Their implementation is crucial for achieving high reliability and robustness characteristics in the power MOSFET and guarantee the lifetime required by the specific application.

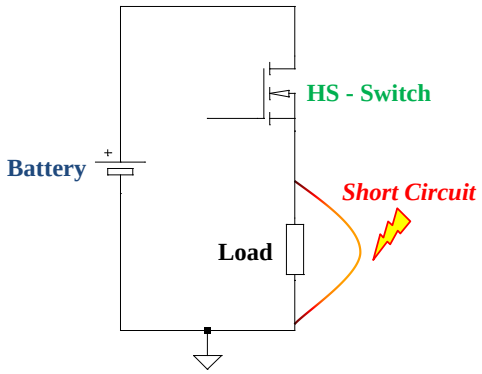

Figure 1.20: Example of a short circuit occurrence across a generic load in the high-side configuration. The full battery voltage drops across the power switch causing a substantially high power dissipation in the semiconductor device.

1.3.2 Critical operating conditions

In this section a brief overview will be given about most common critical operating conditions for the DMOS integrated in the SP device in the automotive environment:

- **Short circuit**

  Within the car electronics, a SP device can be connected to a load by means of two different configurations: i.e. *low-side* and *low-side*. In the former, the power switch is connected between the load and the ground (given by the chassis of the car), while in the latter the power switch is connected between the car battery and the load. Electrical connections to both the ground and the battery are implemented via cable harness. In such electrical configurations, a load can be accidentally short-circuited to the battery voltage or to the ground voltage in the low-side or in the high-side arrangements, respectively. In both cases, the full battery voltage will drop across the power MOSFET. An example of SC occurrence for a high-side configuration has been reported in Figure 1.20. For a conducing device, i.e. in its ON-state, a substantially high current may flow through its substrate causing a high power density transient which may lead the switch to failure when not conveniently limited. Therefore, the logic circuitry of the SP device must be capable of recognizing possible SC events and handling them accordingly. The short circuit duration typically depends on the diagnosis and protection circuits.

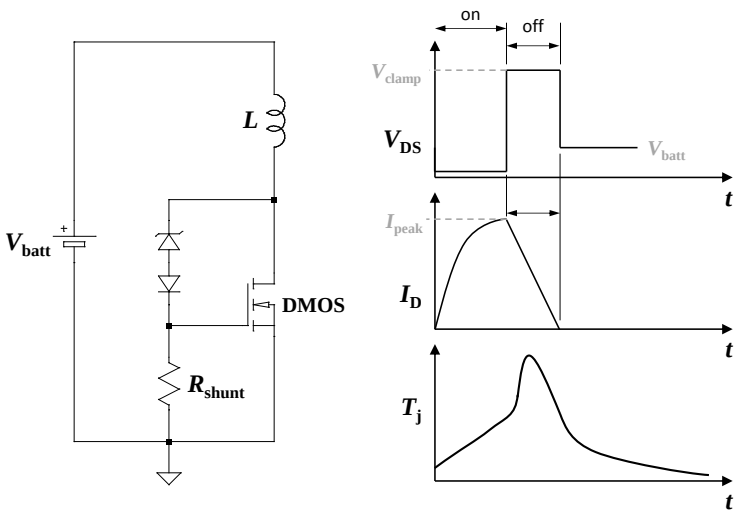

Figure 1.21: Inductive load switching: schematic (left side) and waveforms (right side).

- **Switching-off of an inductive load**

  In Figure 1.21, a typical schematic of a power MOSFET switching-off an inductive load is depicted. In such a configuration, during the turn-off phase, the natural electrical behavior of the inductance tends to oppose the current change by a corresponding increase of its terminal voltage. This electrical behavior results in a transient operation (that may vary in length) in which the power MOSFET must conduct a linear current decrease from I_{peak} to 0 A at high V_{DS} voltage, which in turn determines a substantially high power dissipation. Starting from the characteristic equation of an inductor describing the relation between its voltage and current, the value of the current peak can be determined:

$$I_{\mathrm{peak}} = t_{\mathrm{ON}} \frac{V_{\mathrm{batt}}}{L} \tag{1.30}$$

  where V_{batt} is the battery voltage, t_{ON} is the switching-on time and L the inductance value.

  Concerning the definition of the V_{D} value during the turn-off phase, in voltage-unprotected power MOSFETs, V_{DS} is directly defined by the blocking voltage of the device, namely the breakdown voltage of the

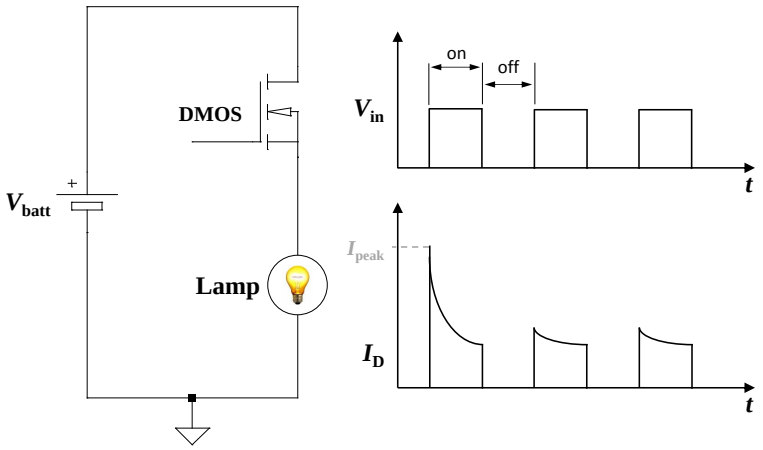

Figure 1.22: In-rush current resulting from the low equivalent resistance associated to the cold filament of a lamp.

body-drain diode. On the other hand, SP devices may feature an overvoltage protection circuit, like the clamping network depicted in Figure 1.21, which "clamps" the potential of the MOSFET drain to a fixed value thanks to one or a series of Z-diodes placed between the gate and the drain terminals. Clamping components are dimensioned in a way that the high current resulting from the magnetic energy discharge flows through the device channel by biasing the gate potential above the V_{th} voltage.

- **In-rush current**

  SP switches are very often used for driving incandescent bulbs of cars' front and back lights. Lamps are not always driven in DC-mode, but often with a sequence of pulses in Pulse Width Modulation (PWM) mode in order to achieve a longer lifetime of the lamp.

  When turning-on the filament of a cold lamp, a high current crosses the power switch for a short transient (few milliseconds) as depicted in Figure 1.22. This phenomenon takes place because the equivalent resistance associated to the cold filament is about 10% of the its equivalent hot-filament resistance and increases for increasing temperature. For

this reason, rather than inductive, at the turn-on lamps typically behave like capacitive loads.

The in-rush current represents a tough operation for the SP switch resulting in a high power density which the power MOSFET has to reliably withstand during the heating phase of the filament.

Other critical operations which are not treated in this work are given by: battery disconnection with inductive load, reverse battery or ESD-discharge [44]. Particularly, in next chapters the attention will be focused on electro-thermal simulations and measurements of short-circuit operation in thermally stable and unstable regimes and switching-off of inductive loads.

Chapter 2

Electro-Thermal Simulation in ANSYS

During the development phase of new device technologies as well as during the verification and/or re-designing phase of already existing ones, the device robustness, and thus the device's SOA margins, must be determined. The usage of dedicated test-chips helps to verify the technology accomplishments in terms of its robustness. This approach generally demands substantial resources and time because it implies the design and the production of dedicated device geometries. Consequently, test chips are normally employed only when a sufficient level of confidence has been achieved in the assessment of proposed design improvements.

A simulation tool can help device designers to estimate with a good level of accuracy the usefulness and effectiveness of the introduced design improvements. This software is of paramount importance because it represents the fastest, cheapest and most easily available method to determine the impact and the goodness of the introduced improvements during the technology/design developments. An electro-thermal simulator is a tool useful for evaluating the device robustness margins by taking into account the electro-thermal interactions normally exhibited by power semiconductor devices, such as the Power MOSFET, the Insulated-Gate Bipolar Transistor (IGBT), the Power BJT, etc..

In this chapter electro-thermal simulations will be introduced. Subsequently, the common methods used to implement the simulation algorithm and the state of the art will be described. The simulation method proposed in this thesis and its implementation in the FEM software ANSYS will be presented and discussed in detail. Finally, the chapter ends with a comparison of simula-

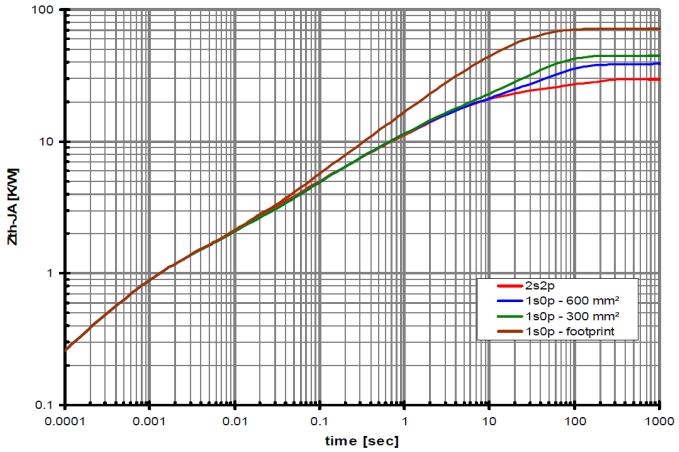

Figure 2.1: Thermal impedance diagram for different board types and cooling area dimensions of a commercial SP device [75].

tions results with experimental measurements performed with a test structure with the aim of validating the electro-thermal simulator.

2.1 Introduction to electro-thermal simulators

2.1.1 Usefulness of electro-thermal simulations

Power MOSFETs for industrial and automotive applications generally have to withstand substantial power dissipation during operation. They have to work reliably even under extreme temperature and electrical stress, such as transient startup or overload conditions Paragraph 1.3.2.

The standard approach used to assess the robustness of a power device is based on the evaluation of maximum allowed voltage, current and junction temperature. For an electrical load pulse of given amplitude and duration, the corresponding temperature rise can be computed based on the thermal impedance diagram reported in the datasheet of the device. As an example Figure 2.1 shows such a diagram for a commercial Smart Power, i.e. *PROFET*^{TM}*BTS5020-2EKA* [75]. Unfortunately, such a thermal impedance diagram is useful to estimate only the mean temperature difference between the junction and the ambient temperature and does not consider the possible power and temperature spatial non-uniformity generally shown by the power device. In fact, under typical overload conditions, a spatially inhomogeneous

power dissipation has been observed to drive these devices into an unstable self-heating effect which may lead the device to failure even when traditional robustness limits are not exceeded, whereas the traditional robustness limits are determined by purely thermal considerations.

Another extremely important aspect is related to the dependence of the drain current of every cell upon the temperature. As discussed already in Paragraph 1.2.3, depending on the local biasing conditions of every device cell, a power MOSFET might show either a positive or a negative α (Drain Current Temperature Coefficient), which in turn has a relevant influence on the overall switching behavior.

In Paragraph 1.2.2 the drain current has been defined as temperature dependent because of the temperature dependencies of the electron mobility in the channel μ_{n} and of the threshold voltage V_{th}. Extending those temperature dependencies to the drain current of a single device cell, one can write for the generic cell j:

$$I_{\mathrm{D}_j} = \begin{cases} \text{if } \; V_{\mathrm{GS}_j} < V_{\mathrm{th}} : \\ 0 \\ \\ \text{if } \; V_{\mathrm{GS}_j} \geq V_{\mathrm{th}} \; \text{ and } \; V_{\mathrm{GS}_j} - V_{\mathrm{th}} \geq V_{\mathrm{DS}_j} : \\ \frac{1}{2}\mu_{\mathrm{n}}(T_j) C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left[2 \left(V_{\mathrm{GS}_j} - V_{\mathrm{th}}(T_j) \right) V_{\mathrm{DS}_j} - V_{\mathrm{DS}_j}^2 \right] \\ \\ \text{if } \; V_{\mathrm{GS}_j} \geq V_{\mathrm{th}} \; \text{ and } \; V_{\mathrm{GS}_j} - V_{\mathrm{th}} < V_{\mathrm{DS}_j} : \\ \frac{1}{2}\mu_{\mathrm{n}}(T_j) C_{\mathrm{OX}} \frac{W_{\mathrm{cell}}}{L_{\mathrm{ch}}} \left[V_{\mathrm{GS}_j} - V_{\mathrm{th}}(T_j) \right]^2 \end{cases} \tag{2.1}$$

$$\mu_{\mathrm{n}}(T_j) = \mu_0 \left(\frac{T_j}{T_0} \right)^{-m} \tag{2.2}$$

$$V_{\mathrm{th}}(T_j) = V_{\mathrm{th}_0} - \phi \left(T_j - T_0 \right) \tag{2.3}$$

The Equation 2.1 defines two V_{GS} ranges where the thermal stability condition can be either satisfied or not, based on the sign of the temperature coefficient α. The stability condition can be achieved whenever the local biasing voltage V_{GS_j} of the elementary cell is greater than voltage value of the temperature compensation point V_{TCP}, as depicted in the temperature dependent transfer characteristics in Figure 2.2.

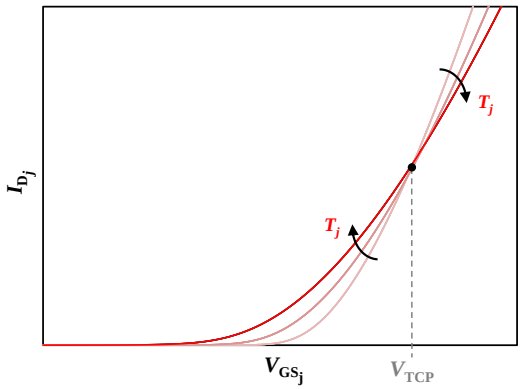

Figure 2.2: Temperature effects on the transfer characteristic of the generic cell of a power MOSFET and definition of the TCP.

Being a power device made of several hundreds of thousands elementary cells connected in a parallel manner, the total drain current in a power MOSFET is the result of all cell current contributions:

$$I_{\mathrm{D}}(T) = \sum\nolimits_j I_j(T_j) \tag{2.4}$$

If a non-uniform temperature distribution exists across the device active area, then every single device cell will be characterized by its own temperature according to the temperature distribution. Therefore, every cell will work on a precise point of its transfer characteristic depending on its temperature and local biasing conditions. The temperature imbalance among cells hence implies a current imbalance (due to Equation 2.1) and consequently a power dissipation imbalance. Finally, the local thermally dissipated power associated to a cell will affect the cell temperature itself. The larger the cell power dissipation, the higher its temperature due to obvious thermal considerations. Depending on the thermal stability condition defined by the device operating point, hotter cells can either increase (thermal instability) or decrease (thermal stability) their currents. The consequent current changes within an elementary cell will affect again the power dissipation and the temperature associated to the cell. Therefore, the described cell behavior consists of a double-verse interaction between the electrical and thermal quantities of the elementary cell.

When considering the R_{ON} parameter in power devices, BJTs typically show a lower value than power MOSFETs allowing better performance in terms of dissipated power. Despite that, they are intrinsically unstable devices

versus temperature, in fact a small temperature imbalance within the device active region leads to a current imbalance among device cells, namely the hotter cells conduct higher current than the colder ones. This phenomenon, called *second breakdown* [43], is due to the positive temperature coefficient of the BJT current gain parameter β and might lead the device to a thermal runaway failure.

The main advantage of power MOSFETs in switching applications is given by their higher switching speed capability when compared with bipolar switches [1, 2, 3]. When the MOSFET was introduced in power electronics, it was supposed to be an intrinsically stable device. Soon it was discovered that its intrinsic stability can be ensured only in certain operating conditions due to a totally unexpected failure mechanism of the device which typically appears in operating conditions characterized by a high voltage and low current. With such biases the device is driven into an unstable region of its transfer characteristic causing its failure when not properly controlled.

We now assume that all the device cells share the same defined biasing condition:

$$V_{\mathrm{GS}_j} = V_{\mathrm{GS}} \quad \forall j = 1 \ldots \mathrm{N} \tag{2.5a}$$

$$V_{\mathrm{DS}_j} = V_{\mathrm{DS}} \quad \forall j = 1 \ldots \mathrm{N} \tag{2.5b}$$

where N is the total number of elementary cells. We will see in the following paragraphs that this is only a simplification and that in general the device cells do not share the same source potential because of the de-biasing effect. Under the conditions of Equation 2.5a and Equation 2.5b, three different cases are possible:

1. All cells have the same temperature (note that this assumption is true at least at the beginning of a transient in stationary conditions). In this case, sharing the same temperature and the same biasing conditions, all the cells carry the same current amount because they all work with the same operating point on the transfer characteristic in Figure 2.2;

2. Temperature differences are present among the cells and the biasing conditions define a negative α. In this case because of the negative feedback between the temperature and the current, a possible temperature inhomogeneity in the device is compensated by a current balancing, which in turn results in a power dissipation balancing and thus a temperature homogenization. This happens because "hotter cells tend to dissipate

lower power" because of the negative feedback between temperature and current.

3. Temperature differences are present among the cells and the biasing conditions define a positive α. In this case because of the positive feedback between the temperature and the current, a possible temperature inhomogeneity in the device is emphasized by a higher current mismatch among the cells, which in turn results in a more inhomogeneous power dissipation and thus a more inhomogeneous temperature distribution. This happens because "hotter cells tend to dissipate an increasing power" because of the positive feedback between temperature and current. This mechanism lead to the emerging of a hot-spot within the device active area localized in the place where the current density is highest (current focusing or current crowding).

The normal device operation in standard applications requires thermally stable biasing conditions. In these cases, the possible emerging of a hot-spot is mainly due to thermal reasons which are directly related to the device geometry and to thermal boundary conditions. However, during critical operations (e.g. short circuit event, turn-on, turn-off, etc.), the device may operate with biasing conditions within the thermally unstable regime. In this case, the formation of the hot-spot associated with the current crowding phenomenon may lead the device to failure because of the thermal runaway effect, similar to the *second breakdown* effect observed in power BJTs [43]. Therefore, in order to evaluate power MOSFET robustness in a given application, the mutual interaction between thermal and electrical behavior which takes place within the power device must be absolutely taken into account. An electro-thermal simulator is a software tool capable to predict the device functioning by correctly simulating such mutual interaction. Consequently, this tool is of paramount importance for the prediction of possible device failures due to the thermal runaway phenomenon in both thermally stable and unstable device behavior.

2.1.2 Mathematical formulation

An electro-thermal simulator must be capable of predicting the temperature distribution within the device in transient operation, given the device geometry and the electrical and thermal material properties of the device structure. Generally, the simulation inputs are the device biasing conditions (which can eventually vary in time) and the ambient temperature at the beginning of the

transient. The outputs are the temperature, the current and the potential distributions within the whole device structure. Therefore, such a software must be capable of evaluating the heat propagation and the electric field distribution in the device geometry and, finally, of coupling accordingly the two fields in order to take into account their mutual interaction. The device geometry defines the domain where both the electrical and thermal field have to be computed.

The way the heat propagates within matter is described by the heat equation. This equation can be derived from *Fourier's law* and from the *law of conservation of energy*. Given a three-dimensional temperature field $T(\boldsymbol{x}, t) = T(x, y, z, t)$ within a domain Ω, the heat equation in a Cartesian coordinate system can be written by means of a Partial Differential Equation (PDE) [76]:

$$\rho(T)c(T)\frac{\partial T(\boldsymbol{x}, t)}{\partial t} - \nabla\tilde{\lambda}(T)\nabla T(\boldsymbol{x}, t) = Q_{\mathrm{V}} \tag{2.6}$$

In Equation 2.6 ρ is the density, c is the specific heat and $\tilde{\lambda}$ is the thermal conductivity tensor matrix. The thermal conductivity may be an anisotropic (or orthotropic) physical quantity, hence it is generally represented with a tensor matrix:

$$\tilde{\lambda} = \begin{bmatrix} \lambda_{xx} & \lambda_{xy} & \lambda_{xz} \\ \lambda_{yx} & \lambda_{yy} & \lambda_{yz} \\ \lambda_{zx} & \lambda_{zy} & \lambda_{zz} \end{bmatrix} \tag{2.7}$$

The reader should notice that in Equation 2.6 ρ, c and $\tilde{\lambda}$ depend on the temperature T, which means that those quantities define *nonlinear* material properties. Finally, the right-hand side of the equation, namely Q_{V}, describes a thermal volumetric heat source included in the domain Ω and is usually called *generation term* of the heat equation.

The heat equation describes the way how the temperature field $T(\boldsymbol{x}, t)$ varies in space and in time. Being a PDE, the definition of an initial condition and a Boundary Condition (BC) are required for determining its solution. The initial condition defines $T(\boldsymbol{x}, t)$ at the time t = 0:

$$T(\boldsymbol{x}, 0) = T_0(\boldsymbol{x}) \tag{2.8}$$

Before describing the possible boundary conditions for the heat equation, it is convenient to introduce the heat flux vector $\boldsymbol{q}$ which is directly defined by Fourier's law:

$$\boldsymbol{q} = -\tilde{\lambda}\nabla T \tag{2.9}$$

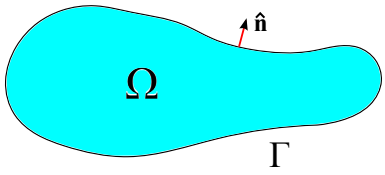

Figure 2.3: Mathematical domain with its boundary curve and the versor normal to the boundary.

For the definitions of boundary conditions, the component of $\boldsymbol{q}$ orthogonal to the domain boundary Γ (also called outward normal heat flux q_{n}) is particularly useful:

$$q_{\mathrm{n}} = \boldsymbol{q} \cdot \hat{\mathbf{n}} = -\tilde{\lambda} \nabla T \cdot \hat{\mathbf{n}} = -\tilde{\lambda} \frac{\partial T}{\partial \hat{\mathbf{n}}} \tag{2.10}$$

where $\hat{\mathbf{n}}$ is the outward versor normal to Γ (see Figure 2.3). The domain Ω may exchange heat with the external environment by means of three different heat transfer mechanisms, i.e. conduction, convection and radiation. Consequently, three kind of formulations can be defined for the boundary condition based on the considered heat exchange mechanism:

1. For a conductive heat exchange, there are two kinds of possible BC definitions. A first one, called Dirichlet boundary condition, defines the temperature values along the domain boundary Γ:

$$T(\boldsymbol{x}) = \overline{T} \quad \forall \boldsymbol{x} \in \Gamma \tag{2.11}$$

   A second condition, called Neumann BC, can be used instead of the Dirichlet BC, by defining the value of the outward normal heat flux in Equation 2.10 :

$$q_{\mathrm{n}}(\boldsymbol{x}) = -\tilde{\lambda} \frac{\partial T}{\partial \hat{\mathbf{n}}} = \overline{\boldsymbol{q}} \quad \forall \boldsymbol{x} \in \Gamma \tag{2.12}$$

   This condition defines the orthogonal heat flux along the boundary. The condition in Equation 2.12 is also called *inhomogeneous* Neumann BC as opposed to *homogeneous* Neumann BC, defined by:

$$q_{\mathrm{n}}(\boldsymbol{x}) = -\tilde{\lambda} \frac{\partial T}{\partial \hat{\mathbf{n}}} = 0 \quad \forall \boldsymbol{x} \in \Gamma \tag{2.13}$$

The BC in Equation 2.13 describes the situation where no heat exchange takes place between Ω and the external environment so that the domain is thermally isolated (adiabatic).

2. For a convective heat exchange between the domain and the external environment, the BC defines the orthogonal heat flux along the boundary as proportional to the temperature difference between the surface temperature and the ambient temperature of the surrounding medium, i.e. $\Delta T = T_\Gamma - T_{\mathrm{medium}}$, by means of a coefficient h called *heat transfer coefficient* or *film coefficient*:

$$\boldsymbol{q}_{\mathbf{n}}(\boldsymbol{x}) = -\tilde{\lambda}\frac{\partial T}{\partial \hat{\mathbf{n}}} = h\Delta T \quad \forall \boldsymbol{x} \in \Gamma \tag{2.14}$$

3. The last boundary condition refers to a radiative heat exchange and defines q_{n} based on the *Stefan-Boltzmann law*:

$$\boldsymbol{q}_{\mathbf{n}}(\boldsymbol{x}) = -\tilde{\lambda}\frac{\partial T}{\partial \hat{\mathbf{n}}} = \epsilon\sigma\left(T_\Gamma^4 - T_{medium}^4\right) \quad \forall \boldsymbol{x} \in \Gamma \tag{2.15}$$

In Equation 2.15 ϵ is the emissivity of the boundary surface (defined as the ratio of the heat emitted by the surface to the heat emitted by a black body at the same temperature) and σ is the Stefan-Boltzmann constant.

Concerning the electrical problem, some considerations are needed for the description of the electric potential within the semiconductor device structure. The differential form of *Faraday's law* is generally written as [77]:

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2.16}$$

which states that the curl of the electric field $\boldsymbol{E}$ is given by the rate of change of the magnetic field $\boldsymbol{B}$ over time. In the following Paragraph 2.3.1 we will see that for the solution of the electro-thermal problem it is sufficient to consider only the stationary case, therefore the time dependencies of the involved electromagnetic quantities can be neglected and Equation 2.16 can be rewritten as:

$$\nabla \times \boldsymbol{E} = 0 \tag{2.17}$$

An important consequence of the Equation 2.17 is that $\boldsymbol{E}$ is irrotational, therefore an electric scalar potential ϕ can be introduced in order to express $\boldsymbol{E}$ as the gradient of ϕ:

$$\boldsymbol{E} \triangleq -\nabla\phi \tag{2.18}$$

In addition to the Equation 2.18, the law of the *continuity of the electric charge* is needed:

$$\nabla \cdot \boldsymbol{J} + \frac{\partial \rho_{\mathrm{e}}}{\partial t} = 0 \tag{2.19}$$

Where ρ_{e} is the density of charge per unit volume per unit time. In the stationary case the current density vector field $\boldsymbol{J}$ is solenoidal, therefore the Equation 2.19 reduces to:

$$\nabla \cdot \boldsymbol{J} = 0 \tag{2.20}$$

The vector $\boldsymbol{J}$ can be related to the electric scalar potential by means of the well known differential form of *Ohm's law*:

$$\boldsymbol{J} = \tilde{\sigma}\boldsymbol{E} \tag{2.21}$$

Where $\tilde{\sigma}$ is the tensor of the electrical conductivity which takes into account possible anisotropies in the considered material:

$$\tilde{\sigma} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{2.22}$$

Substituting Equation 2.21 in Equation 2.20 and subsequently in Equation 2.18, we obtain:

$$\nabla\tilde{\sigma} \cdot \nabla\phi = 0 \tag{2.23}$$

Thus defines the potential distribution within the domain Ω in stationary conditions. The Equation 2.23, as in the case of the heat equation, is a PDE and determining its solution requires the definition of a boundary condition. In this case either a Dirichlet BC or a homogeneous Neumann BC can be defined:

$$\phi = \overline{\phi} \quad \forall \boldsymbol{x} \in \Gamma \tag{2.24}$$

$$\frac{\partial \phi}{\partial \hat{\mathbf{n}}} = 0 \quad \forall \boldsymbol{x} \in \Gamma \tag{2.25}$$

Obviously, no initial condition is required for the solution of Equation 2.23 since the time variable is not involved.

2.1.3 State of the art: simulation approaches and types of coupling

The Equation 2.6 and the Equation 2.23, together with their respective boundary and initial conditions, represent the set of equations that an electro-thermal

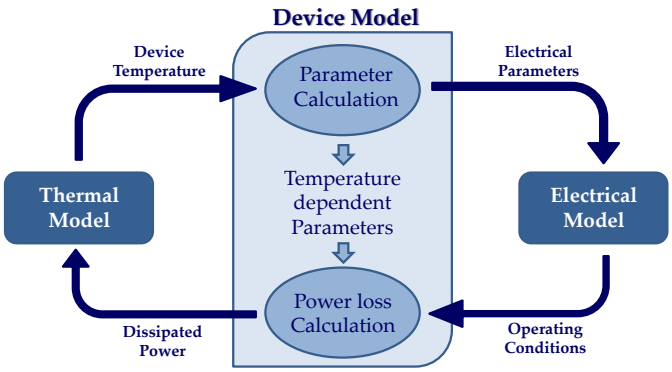

Figure 2.4: Interaction scheme between the device, the thermal and the electrical model implemented in a common electro-thermal simulator [78].

simulator must solve iteratively in order to take into account the mutual interaction between the electrical and the thermal fields. Along with the solution of the two PDEs, the simulator must also solve the Equation 2.1 which describes the electrical behaviour of the elementary cell of the MOSFET. The coupling between the electrical and the thermal field is necessary due to the temperature dependencies of the MOSFET parameters, specifically μ_n and V_{th} defined in Equation 2.2 and Equation 2.3, respectively.

In Figure 2.4 the interaction mechanism between the device, the thermal and the electrical models is schematically depicted. The power dissipated within the device defines the temperature distribution in the device geometry by means of an adequate thermal model. The instantaneous temperature distribution feeds the device model which includes the interaction between the thermal and the electrical behavior. Therefore, the temperature dependent parameters will describe a new electrical representation of the device according to the temperature variation. Consequently, the new electrical model defines a new device operating point accordingly. Finally, the newly computed operating condition defines a new power dissipation value which will be delivered again to the thermal model in order to calculate the new temperature distribution. This loop normally is executed until a convergence condition is satisfied.

The Equation 2.6 and the Equation 2.23 are partial differential equations that must be solved in a generally complex three dimensional geometry and for which no closed form solution exists. For this reason, the solution of the

electro-thermal problem can only be derived using a computational method which considers a numerical approximation of the involved PDEs. Obviously, the adopted numerical method returns only an approximated solution and that solution should be as accurate as possible, i.e. the computed solution must be as close as possible to the real one.

For solving the electro-thermal problem in a power semiconductor device two coupling approaches have been proposed in literature:

- The **relaxation method** (also called *two-step method*) is based on the coupling of two distinct simulators which are dedicated to solving separately the electrical and the thermal problem [26, 27, 28, 29]. The sets of thermal and electrical equations are solved within the two simulation environments in several iteration loops. For example, the electrical problem is commonly solved by means of a SPICE-like simulator while the thermal problem is solved using a finite difference or a finite element software. Typically, the modeling effort required with this approach is not as high as in the case of the direct approach because both simulators are designed to properly solve their respective problems. The main efforts required by the implementation of such a coupling method concern the fulfillment of an adequate data transfer strategy between the two simulators, the synchronization and the strategy adopted for a reliable convergence control. Main advantages are represented by the high accuracy of the simulation results and the simplicity of the modeling process. The main drawback is related to the typically high computational effort required by simulations which implies a substantially long simulation time and the (almost mandatory) usage of high performance hardware, such as a workstation with parallel computing capabilities.

- In the **direct method** the electrical and thermal device behavior are simulated within the same simulation environment since the software tool is capable of directly simulating both coupled fields. Such a tool can be either a SPICE-like circuit simulator, a system simulator (e.g. SABER or Spectre) or a software developed on purpose. This approach requires a prior thermal and electrical modeling of the device. For example the thermal modeling can be done using an analog behavioral language and the electric problem could be described by means of a SPICE net-list. The main advantage of this coupling method is given by the relatively lower computational effort which makes this approach particularly suitable for simulators used within company or industrial environments. The main drawback is represented by the effort required to model accurate

thermal and electrical device behaviour which is a quite complex and time consuming task. In such a simulation approach the main challenge is represented by the thermal modeling of the device. Typically the thermal device behavior can be described using several methods, such as: finite differences, finite element, Fourier series and, most common, thermal ladder networks. The thermal network includes RC lumped parameters which can be determined with either a Foster or a Cauer representation. Such a representation requires the provision of a set of thermal resistances and capacitances, hence besides the method used to solve the thermal problem, further methods are needed to extract model parameters from experimental results. Parameter extraction is generally performed using finite difference, finite element or analytical methods [79, 80].

Most electro-thermal simulators proposed in literature exploit the direct method: Irace et al. developed their simulator entirely in MATLAB solving the electrical and the thermal problem using a forward iterative finite difference time domain scheme [30]. Most simulators based on the direct method exploit a RC network for describing the thermal behaviour of the device [31, 32]. A quite novel approach uses the reduction of the thermal FEM model order by means of a dedicated tool which leads to a lower order equivalent thermal network [81]. With this method the simulation time can be significantly reduced, but the obtained thermal network is often difficult to interpret [33].

In this work is presented a completely new approach to simulate the electro-thermal interaction in a power device by fully exploiting the finite element method. Such an approach could be considered as a sort of *hybrid method* because it uses one single simulator (like in the direct approach) to implement the electro-thermal coupling following a typical relaxation scheme by using two distinct solvers. This feature allows to exploit benefits of both discussed methods and, at the same time, it avoids their respective drawbacks. Specifically, the proposed simulation technique implies two substantial advantages compared to aforementioned coupling methods. First, the possibility to perform simulations by means of just one single software environment avoiding, at the same time, all the programming effort required to implement the numerical method from scratch. Indeed, nowadays commercial software suites, such as ANSYS® or COMSOL Multiphysics®, offer many facilities for solving electrical and thermal problems by FEM analysis. In this sense, the main implementation effort lies in the geometrical modeling of the device, which is anyhow required also by both the relaxation and the direct approach. Sec-

ond, the prior electrical and thermal modeling of the device required by the direct method becomes considerably simplified because it comes down to the definition of device material properties.

2.2 FEM analysis of the electro-thermal problem in Power MOSFETs

2.2.1 The finite element method

A numerical method for the simulation of a physical system is a procedure made of several steps which approximates with "easier" equations one or more PDEs for which no closed form exists. The derived set of equations allows the computation of the solution of the considered physical problem. The procedure typically leads to the definition of a system of (linear and/or nonlinear) algebraic equations. Since the method is based on a mathematical approximation of differential equations, the computed solution is affected by an unavoidable accuracy error: the higher order the introduced approximation, the higher the accuracy of the solution. Nowadays many numerical methods are available for the solution of PDEs, the most common are the boundary element method (BEM), the finite difference method (FDM) and the finite element method (FEM). Among them the FEM method has become the most popular one due to its particularly attractive features, which are: a very high numerical efficiency; the well established theory for the modeling of non linear material properties; the possibility to treat very complex geometries with a reasonably simple modeling effort [82].

The main feature of the FEM approach is the discretization of the mathematical domain by means of Finite Element (FE). The discretization process consists of a graphical subdivision of the computational domain into smaller sections (areas or volumes) called *elements*. For a two dimensional domain either triangles or quadrilaterals or both can be used, while for a three dimensional domain typically tetrahedron, hexahedron and prism are employed. The collection of the finite elements used to represent the computational domain defines the *mesh* of the model and the discretization process is usually called *meshing*. The usage of an effective meshing strategy is a key point for the accuracy of the results obtainable by finite element analysis. Therefore this topic is currently widely studied in literature and most of the commercial FEM software suites provide to users facilities and tools for realizing efficient mesh paths in a totally automatic, semi-automatic or manual manner. Figure 2.5

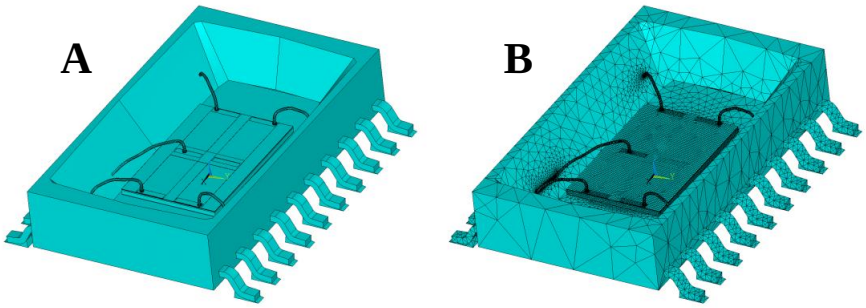

Figure 2.5: Example of a 3-D model of a chip (A) with its respective mesh (B) [83].

shows by way of example the geometrical 3-D model and its respective mesh of a chip with its package.

Due to the wide diffusion that FE method has known in the last decades, many software suites are nowadays available on the market for finite element analysis and Computer-Aided Engineering (CAE) of various kinds of physic problems such as mechanical, acoustic, thermal, fluid dynamics and electromagnetic problems. Currently, the most popular software suites are ANSYS®, COMSOL Multiphysics® and Abaqus FEA. Independently from the used software platform, the solution of a general engineering problem by means of a FEM software suite always requires the definition of following steps:

1. The creation of the 2-D or 3-D geometrical model of the computational domain by means of computer-aided design (CAD) tool;
2. The generation of the model mesh (*meshing*);
3. The assignation of the material properties to the elements of the mesh;
4. The definition of the boundary conditions and loads;
5. The definition of the initial condition (only in cases where a transient analysis has to be performed);
6. The definition of the simulation analysis, i.e. static, transient, harmonic, eigenfrequency analysis, etc.;

7. The definition of the simulation algorithms;
8. The simulation computation;
9. If required, the definition of special operations, such as optimization of design parameters or of objective functions;
10. The retrieval of simulation results, their plotting and analysis.

In the list above the steps 1-5 define the *pre-processing* steps while the step 10 is usually called *post-processing* of the FEM simulation.

2.2.2 Finite element formulation of the electro-thermal problem

In the previous paragraph the FEM simulation approach has been qualitatively described. In this section the finite element mathematical formulation of the electro-thermal problem will be shortly presented starting from the PDEs already derived in the Paragraph 2.1.2. In order to simplify the following analysis, the heat equation (Equation 2.6) in the case of linear and isotropic material properties has been considered:

$$\rho c \frac{\partial T(\boldsymbol{x}, t)}{\partial t} - \lambda \nabla^2 T(\boldsymbol{x}, t) = Q_{\mathrm{V}} \tag{2.26}$$

The Equation 2.26 together with the appropriate initial and boundary conditions presented in Paragraph 2.1.2 represents the *strong formulation* of the thermal problem. The goal of this analysis is to solve this equation, which means finding the temperature field values in space and time $T(\boldsymbol{x}, t)$. For simplicity, associated with the Equation 2.26, we consider a homogeneous Neumann boundary condition:

$$\boldsymbol{q}_{\mathbf{n}} = -\lambda \frac{\partial T}{\partial \hat{\mathbf{n}}} = -\lambda \nabla T \cdot \hat{\mathbf{n}} = 0 \quad \forall \boldsymbol{x} \in \Gamma \tag{2.27}$$

The next step is the introduction of a *test function* ω: a special function which is zero only on the domain boundary Γ:

$$\omega = 0 \quad \forall \boldsymbol{x} \in \Gamma \tag{2.28}$$

The test function ω and the Equation 2.26 must be multiplied and integrated over the computational domain Ω:

$$\int_{\Omega} \omega \left(\rho c \frac{\partial T}{\partial t} \right) \mathrm{d}\Omega - \int_{\Omega} \omega \left(\lambda \nabla^2 T \right) \mathrm{d}\Omega = \int_{\Omega} \omega \left(Q_{\mathrm{V}} \right) \mathrm{d}\Omega \tag{2.29}$$

By applying the integration by parts, the second term of Equation 2.29 can be rewritten as follows:

$$\int_{\Omega} \omega\lambda\nabla^2 T \,\mathrm{d}\Omega = \int_{\Gamma} \omega\lambda\nabla T \cdot \hat{\mathbf{n}} \,\mathrm{d}\Gamma - \int_{\Omega} \lambda\nabla\omega \cdot \nabla T \,\mathrm{d}\Omega \tag{2.30}$$

The reader should notice that the first term on the right-hand side of the equation above is zero due to the BC defined in Equation 2.27. Therefore, by substituting Equation 2.30 into Equation 2.29 we finally obtain the *weak formulation* of the heat equation:

$$\int_{\Omega} \omega \left(\rho c \frac{\partial T}{\partial t}\right) \,\mathrm{d}\Omega + \int_{\Omega} \lambda\nabla\omega \cdot \nabla T \,\mathrm{d}\Omega = \int_{\Omega} \omega Q_{\mathrm{V}} \,\mathrm{d}\Omega \tag{2.31}$$

The weak formulation, also called *variational formulation* [84], is an equivalent formulation of the strong formulation and it can be proven that this formulation is mathematically equivalent to the strong one [85]. The next step of the finite element formulation is the discretization of the computational domain Ω into an integer number n_e of small cells, so called finite elements, by means of an appropriate meshing strategy as already discussed in the previous paragraph. While meshing the computational domain, the elements have to satisfy two main conditions: the domain should be covered by elements as much as possible:

$$\bigcup_{m=1}^{n_{\mathrm{e}}} e_m \subseteq \Omega \tag{2.32}$$

and no intersections must exist between them:

$$\bigcap_{m=1}^{n_{\mathrm{e}}} e_m = \emptyset \tag{2.33}$$

An example which satisfies both above conditions is displayed in Figure 2.6, where a two dimensional domain has been meshed with triangular elements only. The intersections of the grid lines define the *nodes* of the mesh. The nodes are the sites where approximate solutions for the unknowns, i.e. the temperature field values, are computed by means of the FE method.

The core of FEM is the introduction of piecewise polynomial functions (so called *basis* or *shape* functions N) used for determining the approximated solutions at the nodes. The shape function is a special kind of function defined with a local support, namely the function is zero everywhere except within

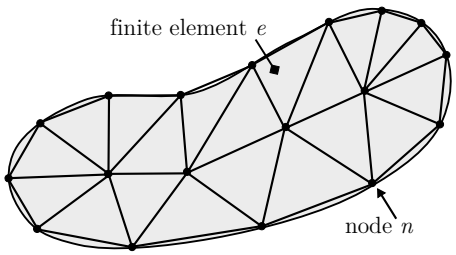

Figure 2.6: Example of a 2-D domain meshed with triangular elements (courtesy of Helmut Köck).

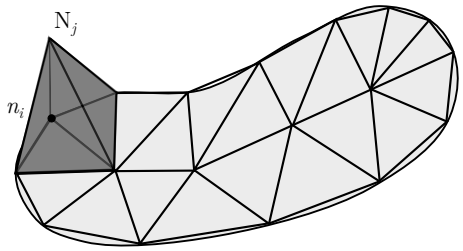

Figure 2.7: Example of linear basis functions in the 2-D mesh of Figure 2.6 (courtesy of Helmut Köck).

its "local domain". Therefore, the generic basis function N_j associated to the generic element e with nodes n_i must fulfill the following condition:

$$N_j(n_i) = \delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \tag{2.34}$$

By way of example in Figure 2.7 linear (first order) basis functions have been applied to the elements set depicted in Figure 2.6. Higher order (second, third, etc.), such as quadratic or cubic, basis functions generally allow higher solution accuracy at the nodes sites, but at the same time require a higher computational effort, hence a trade off between accuracy and computational time must always be considered when using FEM [84]. The introduction of the basis function leads to the following approximation for both the unknown T and the test

function ω:

$$T(t) \approx T^h(t) = \sum_{a=1}^{n_{\text{eq}}} N_a(\boldsymbol{x}) T_a(t) + N_e(\boldsymbol{x}) T_e \tag{2.35a}$$

$$\omega \approx \omega^h = \sum_{b=1}^{n_{\text{eq}}} N_b(\boldsymbol{x}) \omega_b \tag{2.35b}$$

where N_a, N_e and N_b are the basis function of T and ω respectively, and n_{eq} represents the number of equations which is defined by the number of nodes where the unknowns have to be computed. The Equation 2.35a and Equation 2.35b represent the so called *ansatz* of the finite element formulation. That ansatz must be now substituted in the Equation 2.31:

$$\begin{aligned}
&\int_\Omega \rho c \left(\sum_{b=1}^{n_{\text{eq}}} N_b(\boldsymbol{x}) \omega_b \right) \frac{\partial}{\partial t} \left(\sum_{a=1}^{n_{\text{eq}}} N_a(\boldsymbol{x}) T_a(t) + N_e(\boldsymbol{x}) T_e \right) \mathrm{d}\Omega \\
&+ \int_\Omega \lambda \nabla \left(\sum_{b=1}^{n_{\text{eq}}} N_b(\boldsymbol{x}) \omega_b \right) \cdot \nabla \left(\sum_{a=1}^{n_{\text{eq}}} N_a(\boldsymbol{x}) T_a(t) + N_e(\boldsymbol{x}) T_e \right) \mathrm{d}\Omega \\
&- \int_\Omega Q_{\text{V}} \left(\sum_{b=1}^{n_{\text{eq}}} N_b(\boldsymbol{x}) \omega_b \right) \mathrm{d}\Omega = 0
\end{aligned} \tag{2.36}$$

The above expression can be rewritten considering that the spatial derivatives act only on the basis functions (T and ω have already been discretized in space) and that integrals and summations are interchangeable, hence obtaining the *semidiscrete Garlerkin formulation*:

$$\begin{aligned}
\sum_{b=1}^{n_{\text{eq}}} \omega_b \Bigg(\sum_{a=1}^{n_{\text{eq}}} &\int_\Omega \rho c \big(N_a(\boldsymbol{x}) + N_e(\boldsymbol{x}) T_e \big) N_b(\boldsymbol{x}) \, \mathrm{d}\Omega \frac{\partial}{\partial t} T_a(t) \\
&\int_\Omega -\lambda \nabla \big(N_a(\boldsymbol{x}) + N_e(\boldsymbol{x}) T_e \big) \cdot \nabla N_b(\boldsymbol{x}) \, \mathrm{d}\Omega T_a(t) \\
&- \int_\Omega Q_{\text{V}} N_b(\boldsymbol{x}) \, \mathrm{d}\Omega \Bigg) = 0
\end{aligned} \tag{2.37}$$

which can be rewritten in the following matrix form:

$$\mathbf{C}_T \dot{\boldsymbol{T}}(t) + \mathbf{K}_T \boldsymbol{T}(t) = \boldsymbol{f}_T(t) \tag{2.38}$$

where $\boldsymbol{T}(t)$ is the vector of unknown temperatures and $\dot{\boldsymbol{T}}(t)$ is its time derivative, $\mathbf{C}_T$ (with elements C_{ab}) and $\mathbf{K}_T$ (with elements K_{ab}) are the capacity and the conductivity matrices respectively:

$$K_{ab} = \int_\Omega \lambda \nabla N_a(\boldsymbol{x}) \cdot \nabla N_b(\boldsymbol{x}) \, \mathrm{d}\Omega, \qquad 1 \leq a, b \leq n_{\mathrm{eq}} \tag{2.39a}$$

$$C_{ab} = \int_\Omega \rho c N_a(\boldsymbol{x}) N_b(\boldsymbol{x}) \, \mathrm{d}\Omega \qquad 1 \leq a, b \leq n_{\mathrm{eq}} \tag{2.39b}$$

finally $\boldsymbol{f}_T(t)$, with elements f_{be}, is the corresponding right-hand side (usually called *load vector*):

$$f_{be} = \int_\Omega Q_{\mathrm{V}} N_b(\boldsymbol{x}) \, \mathrm{d}\Omega + \int_\Omega N_e(\boldsymbol{x}) T_e \, \mathrm{d}\Omega \quad 1 \leq b \leq n_{\mathrm{eq}}, 1 \leq e \leq n_{\mathrm{bc}} \tag{2.40}$$

The Garlerkin formulation is defined as semidiscrete due to the fact that the discretization process has been accomplished in space but not yet in time. The time discretization process is based on the subdivision of the time interval into M time steps:

$$[0, t_{\mathrm{last}}] = \bigcup_{n=1}^{M} [t_{n-1}, t_n] \tag{2.41}$$

Here a constant time step size has been assumed for simplicity:

$$\Delta t = t_n - t_{n-1} = \frac{t_{last}}{M} \tag{2.42}$$

In many practical cases an *adaptative time step* is also used. With this approach a variable Δt is defined during the solution computation at every iteration based on the level of accuracy desired for the unknowns. This algorithm generally results in a lower computational effort when determining the value of the unknowns especially in cases where the time derivative of the considered field is small, namely when field changes are slow.

Using a *forward time difference* scheme, the time derivative of the temperature can be approximated as follow:

$$\dot{\boldsymbol{T}}(t) \approx \frac{\boldsymbol{T}(t_{n+1}) - \boldsymbol{T}(t_n)}{\Delta t} = \frac{\boldsymbol{T}_{n+1} - \boldsymbol{T}_n}{\Delta t} \tag{2.43}$$

Using the approximation scheme above, we can finally approximate $\dot{\boldsymbol{T}}(t)$ in

the Equation 2.38 with the help of a *time integration parameter* γ_P:

$$\mathbf{C}_T \frac{\boldsymbol{T}(t_{n+1}) - \boldsymbol{T}(t_n)}{\Delta t} + \gamma_P \mathbf{K}_T \boldsymbol{T}(t_{n+1}) + (1-\gamma_P)\,\mathbf{K}_T \boldsymbol{T}(t_n) = \gamma_P \boldsymbol{f}_T(t_{n+1}) + (1-\gamma_P)\,\boldsymbol{f}_T(t_n) \tag{2.44}$$

The value of γ_P allows the usage of different time discretization schemes: $\gamma_P = 0$ defines a forward difference discretization (also called *forward Euler*); $\gamma_P = 0.5$ defines the trapezoidal rule (also known as *Crank-Nicolson rule*) generally achieving a higher accuracy; finally $\gamma_P = 1$ defines a backward difference discretization (or *backward Euler*) [82, 86].

Having seen the finite element formulation for the thermal problem, the formulation for the electrical problem will be shortly presented below under simplified conditions, just by using analogous steps. The starting point is the Equation 2.23 in the case of linear and isotropic material properties:

$$\nabla\sigma \cdot \nabla\phi = 0 \tag{2.45}$$

With the homogeneous Neumann boundary condition:

$$\frac{\partial\phi}{\partial\hat{\mathbf{n}}} = 0 \quad \forall \boldsymbol{x} \in \Gamma \tag{2.46}$$

The Equation 2.45 represents the strong formulation which must be multiplied with a test function ω:

$$\int_\Omega \omega\,(\nabla\sigma \cdot \nabla\phi)\;\mathrm{d}\Omega = 0 \tag{2.47}$$

Integrating by parts, Equation 2.47 becomes:

$$\int_\Omega \sigma\nabla\omega \cdot \nabla\phi\,\mathrm{d}\Omega - \int_\Gamma \sigma\omega\frac{\partial\phi}{\partial\boldsymbol{n}}\,\mathrm{d}\Gamma = 0 \tag{2.48}$$

Due to the BC defined in the Equation 2.46, we finally obtain the following weak formulation:

$$\int_\Omega \sigma\nabla\omega \cdot \nabla\phi\,\mathrm{d}\Omega = 0 \tag{2.49}$$

Using the ansatz:

$$\phi \approx \phi^h = \sum_{a=1}^{n_{\mathrm{eq}}} N_a(\boldsymbol{x})\phi_a + N_e(\boldsymbol{x})\phi_e \tag{2.50a}$$

$$\omega \approx \omega^h = \sum_{b=1}^{n_{\mathrm{eq}}} N_b(\boldsymbol{x})\omega_b \tag{2.50b}$$

into Equation 2.49:

$$\int_\Omega \sigma\nabla \left(\sum_{b=1}^{n_{\mathrm{eq}}} N_b(\boldsymbol{x})\omega_b\right) \cdot \nabla \left(\sum_{a=1}^{n_{\mathrm{eq}}} N_a(\boldsymbol{x})\phi_a + N_e(\boldsymbol{x})\phi_e\right) \mathrm{d}\Omega = 0 \tag{2.51}$$

and rearranging the terms:

$$\sum_{b=1}^{n_{\mathrm{eq}}} \omega_b \left(\sum_{a=1}^{n_{\mathrm{eq}}} \phi_a \int_\Omega \sigma\nabla \left(N_a(\boldsymbol{x}) + N_e(\boldsymbol{x})\phi_e\right) \cdot \nabla \left(N_b(\boldsymbol{x})\right) \mathrm{d}\Omega\right) = 0 \tag{2.52}$$

we finally obtain the matrix notation for the electrostatic equation:

$$\mathbf{K}_\phi \boldsymbol{\phi} = \boldsymbol{f}_\phi \tag{2.53}$$

Where the elements of the *stiffness matrix* $\mathbf{K}_\phi$ and of the load vector $\boldsymbol{f}_\phi$ are defined as follow:

$$K_{ab} = \int_\Omega \sigma\nabla N_a(\boldsymbol{x}) \cdot \nabla N_b(\boldsymbol{x}) \,\mathrm{d}\Omega \qquad 1 \leq a, b \leq n_{\mathrm{eq}} \tag{2.54a}$$

$$f_e = \int_\Omega N_e(\boldsymbol{x})\phi_e \,\mathrm{d}\Omega \qquad 1 \leq e \leq n_{\mathrm{bc}} \tag{2.54b}$$

The finite element formulation seen here allows the numerical computation of the electric scalar potential ϕ in the whole domain at the nodes sites. Based on the knowledge of ϕ, the electrical field can also be determined in the following way:

$$\boldsymbol{E} = -\nabla\phi$$

Consequently, the current density distribution in the computational domain can be derived by means of the differential Ohm's law:

$$\boldsymbol{J} = \sigma\boldsymbol{E}$$

where σ is the electrical conductivity.

Up to now the finite element formulations have been discussed separately for both the thermal and the electrical problems. The approach to consistently couple the two fields together within the same finite element formulation is based on the observation that the electrical power dissipated in the domain Ω is the result of the energy dissipated due to the Joule effect, i.e. the electric field together with the associated electrical conductivity defines a current which generates heat:

$$Q_{\text{el}} = \boldsymbol{E} \cdot \boldsymbol{J} \tag{2.55}$$

Therefore it is consistent to include the electrically generated heat in the right-hand side term of the Equation 2.38 (i.e. the load vector), namely as a heat generation term. Finally, the Equation 2.38 and the Equation 2.53 can be solved together within the same system of algebraic equations:

$$\begin{bmatrix} \mathbf{C}_T & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{T}} \\ \dot{\boldsymbol{\phi}} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_T & 0 \\ 0 & \mathbf{K}_\phi \end{bmatrix} \begin{bmatrix} \boldsymbol{T} \\ \boldsymbol{\phi} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_{\text{th}} \\ \boldsymbol{f}_{\text{el}} \end{bmatrix} \tag{2.56}$$

Where the therm $\boldsymbol{f}_{\text{th}}$ takes into account not only an eventual heat generation term Q_{V} as in the Equation 2.40 but also the Joule heat term Q_{el} in Equation 2.55 [87].

The Equation 2.56 is actually the algebraic system that ANSYS software solves whenever electro-thermal simulations with Joule coupling have to be performed [88]. This kind of approach is very useful every time the power dissipated by resistive materials has to be taken into account in the thermal problem [89, 83, 22, 90, 91]. Unfortunately, such a formulation is still insufficient for simulating the power MOSFET operation due to the nonlinearities introduced by the device, specifically the dependence of the drain current upon the temperature and the local biasing voltage.

2.3 Implementation in ANSYS

2.3.1 Algorithm fundamentals

Due to the electro-thermal coupling exhibited by a semiconductor power device, the finite element formulation shown in Paragraph 2.2.2 cannot be used as it is for performing transient simulations. In more detail the main issue is given by the fact that every single generic cell *j* of the device act as an elementary transistor where the current level is determined by both the biasing voltages and the temperature of the cell itself as reported in Equation 2.1, Equation 2.2

and Equation 2.3. Therefore, a method to let ANSYS take into account the drain current temperature dependence of every cell has been implemented by customizing the software by means of its ANSYS® Parametric Design Language (APDL) language, namely a powerful programming language for optimizing and customizing the FEM workflow in ANSYS [92]. From the modeling point of view a power MOSFET can be seen as a generally thick silicon substrate "covered" on the top with a substantially thinner and more complex layer with different silicon properties, i.e. the epitaxial layer of the device. A power MOSFET is usually a *vertical* device, which means that the current flows from the bottom (drain) to the top (source) of the silicon block, contrary to the standard small-signal MOSFET, which is a *planar* device. In this very simplified geometric model of the device, the electrical resistances as well as the thermal impedance of the silicon substrate depend upon the temperature. This means that in the FE formulation, both the thermal (thermal conductivity λ and specific heat c) and electrical material properties (electrical resistivity ρ) of the silicon substrate are nonlinear:

$$\lambda_{\text{sub}} = \lambda(T)$$
$$c_{\text{sub}} = c(T)$$

$$\rho_{\text{sub}} = \rho(T)$$

In the above expressions the eventual temperature dependence of the electrical capacity associated to the substrate has been intentionally neglected because the simulation approach proposed in this thesis neglects capacitive effects. The reason of this choice will be clarified in the following of this paragraph. Therefore, the substrate layer is described by a pure ohmic model where the current flow generates heat due to the Joule effect, hence this behavior can be easily simulated by means of the standard electro-thermal simulation concept already implemented in ANSYS.

Concerning the device epitaxial layer, the modeling becomes more complicated because it generally includes three main resistive contributions: the channel resistance, the accumulation resistance (which assumes relevant values only for the VD-MOSFET and can be reasonably neglected for a Trench MOSFET) and the drift resistance [54]. The accumulation resistance and the drift resistance can be modeled in the same manner seen for the substrate resistance, i.e. by taking into account the eventual temperature dependencies of the electrical and thermal material properties associated to the corresponding elements in the FE formulation. In particular, the dependence of the electrical

resistivity upon the carrier mobility in a semiconductor is given as [43]:

$$\rho = \frac{1}{\mathrm{q} N \mu(T)} \tag{2.57}$$

where q is the elementary charge, N is the doping level and μ_{n} is the carrier mobility, which is a temperature dependent parameter. Therefore the values of ρ as a function of the temperature should be assessed using an accurate temperature dependent mobility model. In this work, the mobility model proposed by Reggiani et al. [93] has been used for the assessment of the electrical resistivities of both the substrate and the drift zone of the device, while average doping levels of the substrate silicon and of the drift silicon respectively, have been used when evaluating both the mobility and the Equation 2.57. Reggiani et al. experimentally demonstrated that their model returns accurate results up to 700 K.

Being the channel a voltage and temperature controlled resistance, its equivalent resistivity cannot be modeled just by using a purely ohmic approach. In order to correctly take into account the thermal and electrical behavior, the local electrical material properties of every channel element must be tuned at every iteration according to the local biasing condition of the element and to the element temperature. The equivalent electrical resistivity of a generic element j within the channel layer can be expressed using the second Ohm's law as follows:

$$\rho_{\mathrm{ch}_j} = \frac{A_{\mathrm{j}}}{d_{\mathrm{j}}} \frac{\Delta V_{\mathrm{ch}_j}}{I_{\mathrm{ch}_j}(T_j, V_{\mathrm{G}_j})} \tag{2.58}$$

Where ρ_{ch_j} is the electrical resistivity in $\Omega\,$m in the vertical direction (according to the current flow direction), A_j is the area of the element j in m^2 and d_j its thickness in m (see Figure 2.8). ΔV_{ch_j} is the vertical voltage drop in V across the element and I_{ch_j} is the drain current in A which is determined by the element temperature T_j (in K) and its gate potential (in V). In Equation 2.58 the value of the current associated to the element j can be determined by using:

1. an analytical model, for instance the one described by Equation 2.1 with the temperature dependencies of the channel mobility μ_{n} (Equation 2.2) and of the threshold voltage V_{th} (Equation 2.3) according to the temperature value of the element T_j;

2. a table model, i.e. a table where every combination of the triad $(T_{\mathrm{cell}}, V_{\mathrm{GS}_{\mathrm{cell}}}, V_{\mathrm{DS}_{\mathrm{cell}}})$ determines a cell current value. Being a discrete

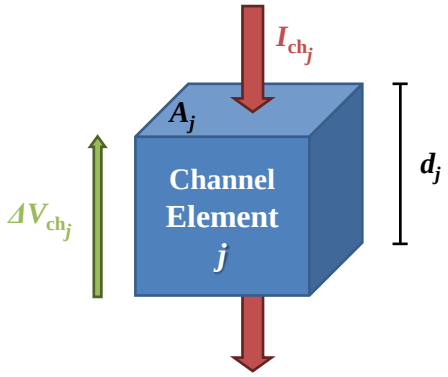

Figure 2.8: A generic channel element j: detail of its current I_{ch_j}, voltage drop ΔV_{ch_j}, element thickness d_j and element area A_j.

model, the usage of the table typically implies the needing of an interpolation algorithm for the assessment of the cell current. Generally, a table model can be obtained either by transfer characteristics measurements or by TCAD device simulations.

The way to deal with such a preemptive material property definition, i.e. ρ_{ch_j}, is to perform a purely electrical finite element steady state simulation just before the electro-thermal simulation, so that the potential distribution can be determined within the device allowing the computation of ΔV_{ch_j} terms for every element j in the channel layer. Once the electrical simulation is finished, the new resistivity values can be defined for the channel elements and the electro-thermal simulation can be started. The electrical steady state simulation is implemented in ANSYS by means of the FE formulation defined with the Equation 2.53. Contrary to the electro-thermal simulation, the prior electrical simulation is not a transient analysis. This choice comes from the observation that time constants associated to the thermal evolution in a power device are generally sufficiently larger than time constants associated to the electrical one, hence the capacitive contributions can be neglected and only the resistive contributions to the electrical impedances must be taken into account. Therefore, a purely steady state simulation is sufficient for the determination of the electrical quantities within the electro-thermal simulation scheme, because the whole dynamic behavior of the power device in a simulated transient event, e.g. a SC event, is dominated by the relatively slow thermal regime.

According to what has been explained above, in order to simulate a generic

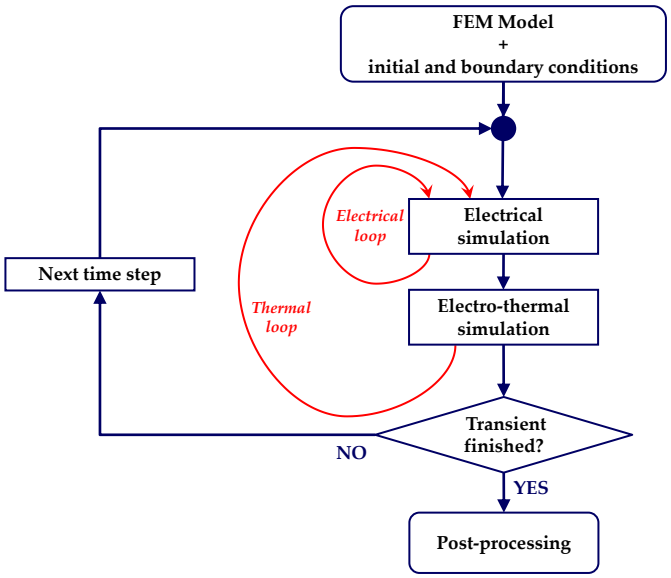

Figure 2.9: Electro-thermal transient simulation algorithm implemented in ANSYS® using APDL.

transient event, the algorithm depicted in Figure 2.9 has been implemented in ANSYS classic using the APDL programming language. Given the FE model of the power MOSFET, the thermal and electrical boundary conditions and the thermal initial condition (details about these conditions will be presented in the next paragraph), the procedure is based on the iterations of two nested cycles [34, 35]:

- The first cycle determines the potential distribution in the device model by means of steady state simulations only. A single electrical simulation would not be enough for determining the potential distribution due to the dependencies of the resistivity values associated to channel elements upon the local biasing conditions of the elements. Indeed, after the first iteration, the Equation 2.58 must be re-computed for every element of the channel zone, subsequently a new electrical simulation can be run again. The iterations last until a convergence criterion for the total computed current of the device is satisfied. Furthermore, such kind of implementation allows also to take into account a possible de-biasing effect due to the finite thermal conductivity of the source metallization

[56, 57, 58]. This phenomenon has generally a strong impact on the electro-thermal behavior of large area power devices where the lateral ohimic voltage drops on the source metal layer are relevant;

- Within the second cycle the temperature and the potential distribution in the device model are computed by means of transient electro-thermal simulations. Also in this case several electro-thermal iterations are needed due to the dependencies of the resistivity values associated to channel elements upon the local biasing conditions and upon the local temperature of the elements. Indeed, a change of the local temperature and of the local biasing voltage within the generic channel element j implies the redefinition of its equivalent electrical resistivity according to Equation 2.58.

The stopping criterion for both the electrical and the electro-thermal cycles is given by:

$$|I_i - I_{i-1}| \leq I_{\text{tol}} \tag{2.59}$$

where I_i and I_{i-1} is the value of the total device current at the iteration i and $i-1$ respectively, and I is the summation of all current contributions of the channel elements, namely:

$$I = \sum\nolimits_j I_j \tag{2.60}$$

The value of I_{tol} in Equation 2.59 can be set based on the desired accuracy for the computation of the device current.

The two nested iterations shown in Figure 2.9 are performed for every time step of the transient simulation. Once all the computations have been performed for all the time steps, a post-processing phase has to be executed for retrieving the results of interest in the simulation.

2.3.2 Initial and boundary conditions

In this section, some considerations will be presented concerning the initial and boundary conditions to be set before performing an electro-thermal simulation.

As already discussed, the initial condition must be defined only for the thermal problem, which means defining the initial temperature of every node. Typically, when performing an electro-thermal simulation of a power device, the thermal initial condition is determined by setting all the nodes in the FE model at the ambient temperature T_{ambient}:

$$T_n(t=0) = T_{\text{ambient}} \quad \forall n \in N \tag{2.61}$$

where n represents the node number and N the total number of nodes in the FE model.

When performing simulations of short pulses (few microseconds), the usually large thermal capacity associated to the lead-frame can be taken into account assuming that no thermal flux exists within the lead-frame geometry. Therefore it is reasonable to assume that the lead frame has constantly the same temperature during the considered transient. This assumption allows a big computational benefit because the FE model does not necessarily include the lead-frame geometry, hence significantly reducing the total number of elements. In all these cases a Dirichlet boundary condition must be applied to the bottom surface of the silicon substrate:

$$T_n = \tilde{T} \quad \forall n \in N_{\text{sub}_{\text{bottom}}} \tag{2.62}$$

In the above condition, normally $\tilde{T}$ is the temperature of the lead-frame, namely the ambient temperature. Alternatively, when the lead frame is also taken into account in the FE model, the Dirichlet boundary applies to its bottom surface:

$$T_n = \tilde{T} \quad \forall n \in N_{\text{lead-frame}_{\text{bottom}}} \tag{2.63}$$

The large thermal resistivity value associated to the package mould compound also implies a relevant computational benefit in case of simulations of short pulses. In fact, in these cases we can assume that no heat exchange takes place between the device model boundaries and the mold compound. Hence, the finite element model can be built neglecting the geometry of the package and all the remaining model boundary surfaces have to be set as adiabatic, i.e. thermally isolated. Therefore, a homogeneous Neumann BC must be applied to those surfaces.

Finally, the electrical boundary conditions are normally defined by the values of the source and the drain potentials outside the device FE model. Typical conditions which have been used in the simulations shown in this thesis are the following:

- Set all the nodes belonging to the top surfaces of the bond wires to the source potential V_{S}:

$$\phi_n = V_{\text{S}} \quad \forall n \in N_{\text{wire}_{\text{top}}} \tag{2.64}$$

- Set all the nodes belonging to the bottom surface of the silicon substrate to drain potential V_{D}:

$$\phi_n = V_{\text{D}} \quad \forall n \in N_{\text{sub}_{\text{bottom}}} \tag{2.65}$$

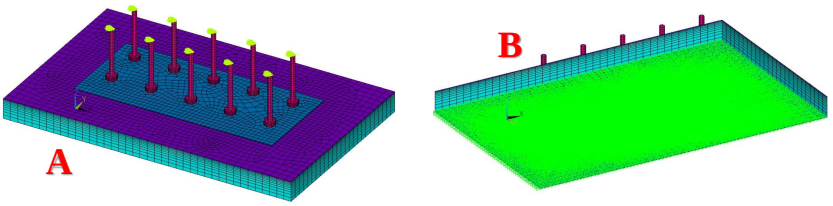

Figure 2.10: Example of electrical BCs (green color): source potential set at the top of the bond wires (A) and drain potential set at the bottom of the silicon substrate (B).

or alternatively, set to drain potential all the nodes belonging to the bottom surface of the lead-frame:

$$\phi_n = V_{\mathrm{D}} \quad \forall n \in N_{\text{lead-frame}_{\text{bottom}}} \tag{2.66}$$

Figure 2.10 shows an example in ANSYS.

As in the thermal case, an electrically homogeneous Neumann BC (electrical isolation) must be applied to all the remaining external surfaces of the FE model due to the very high electrical resistivity of the package mould compound.

In the electro-thermal simulation algorithm presented in Figure 2.9, the electrical boundary conditions define the input waveforms. In fact, the waveforms can be seen as time varying electrical BCs:

$$V_{\mathrm{DS}}(t_k) = V_{\mathrm{DS}_k} = V_{\mathrm{D}}(t_k) - V_{\mathrm{S}}(t_k) \tag{2.67}$$

$$V_{\mathrm{GS}}(t_k) = V_{\mathrm{GS}_k} = V_{\mathrm{G}}(t_k) - V_{\mathrm{S}}(t_k) \tag{2.68}$$

Where t_k represents the time step. Therefore, at the beginning of every simulation, the table $[t_k, V_{\mathrm{DS}_k}]$ and the table $[t_k, V_{\mathrm{GS}_k}]$ must be provided. These waveforms define the biasing conditions of the power MOSFET in the discretized time. When performing an electro-thermal simulation with these waveforms, the simulation outputs consist of the temperature field, the potential and the current density distributions in the device FE model at every time step:

$$\left\{ \begin{array}{c} V_{\mathrm{DS}_k} \\ V_{\mathrm{GS}_k} \end{array} \right. \quad \Rightarrow \quad \begin{array}{l} I_{\mathrm{D}}(t_k) \\ \phi(\boldsymbol{x}, t_k) \\ T(\boldsymbol{x}, t_k) \end{array}$$

In the above scheme I_{D} represents the total device current and this quantity can always be easily computed since the current density distribution is known.

The algorithm presented here is suitable for predicting the electrical and thermal behavior of power semiconductor switches and is particularly convenient for simulating voltage-controlled devices, such as power MOSFETs and IGBTs.

2.3.3 Simulations of Smart Power switches

A SP switch is a chip where one or possibly more power MOSFETs are integrated together with several kinds of circuits (called *logic*) within the same silicon substrate. The logic part comprises the circuitry for protection, diagnosis and driving of the power MOSFET (details about the SP device can be found in Paragraph 1.3.1). Such a kind of switch, besides the drain and source pins, includes an input pin which is usually driven by means of a logic signal suitable to control the ON/OFF power switch status through the integrated gate drivers of the MOSFET. Furthermore, during typical critical operating conditions in automotive applications (such as short circuit, battery disconnection or reverse current, in-rush current, etc.) the device gate is driven by internal protection circuits, so that the DMOS can reliably withstand the electrical stress condition avoiding an eventual failure. Therefore, a SP device cannot be directly driven through its MOSFET gate because no pin is provided for that node.

The simulation algorithm presented in Paragraph 2.3.1 and Paragraph 2.3.2 is suitable for simulating power MOSFETs but not SP switches, because it requires the definition of the gate-source waveform V_{GS} as an input simulation parameter. In this paragraph an extension of the algorithm for allowing simulations of SP switches will be presented. Regarding critical device operation defined by the external electrical environment, the electro-thermal simulation inputs are now V_{DS} and I_{D} waveforms, while no V_{GS} waveform is available anymore.

The algorithm presented in Figure 2.9 must be modified only with regard to the electrical cycles. At the generic time step t_k the electrical boundary conditions are defined by the source V_{S} and the drain potential V_{D} values through the assigned waveforms, moreover the device current I_{D} is known. In this case the electro-thermal problem consists of finding the current and the temperature distributions within the channel elements set that equals the assigned input current, i.e. the imposed current I_{imp}, to the sum of the current contributions of every channel element, called computed current I_{comp}. In other words, the

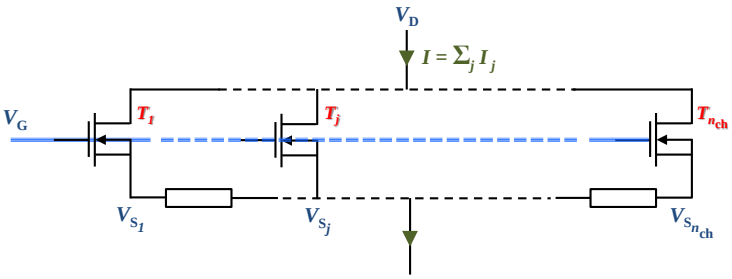

Figure 2.11: Schematic representing the typical parallel configuration of elementary cells in a power MOSFET. Every cell features its own temperature and source potential and shares with all other cells the same gate and drain potential. Finally, the total device current is given by the sum of every single cell current contribution.

simulator, by means of electrical iterations, must minimize the following goal function at every time step t_k:

$$|f_I| = |I_{\text{imp}} - I_{\text{comp}}| \tag{2.69}$$

where I_{comp} is defined in Equation 2.60 and $I_{\text{imp}} = I_{\text{D}_k}$.

During the electrical cycles the resistivity of every channel element must be determined according to Equation 2.58. The element resistivity depends upon the value of the element current I_j and the current contribution of every channel element depends on the local temperature and the local potential according to the Equation 2.1. Unfortunately, while V_{S_j} can be determined within the electrical iterations starting from the boundary condition V_{s}, the gate potential V_{G_j} is now an unknown being not a simulation input anymore. A fundamental aspect of elementary cells composing a power device is that in parallel connection they share the same gate potential as depicted in the schematic in Figure 2.11. Since a channel element includes a certain number of elementary cells, we can also assume that channel elements share the same gate potential value, namely:

$$V_{\text{G}_j} = V_{\text{G}} \quad \forall j \tag{2.70}$$

Therefore, the electrical iterations must solve the Equation 2.69 by properly tuning the V_{G} parameter. This means finding at every time step the gate potential which returns the imposed current I_{imp} according to the current temperature distribution in the channel element layer. As the Equation 2.58 is

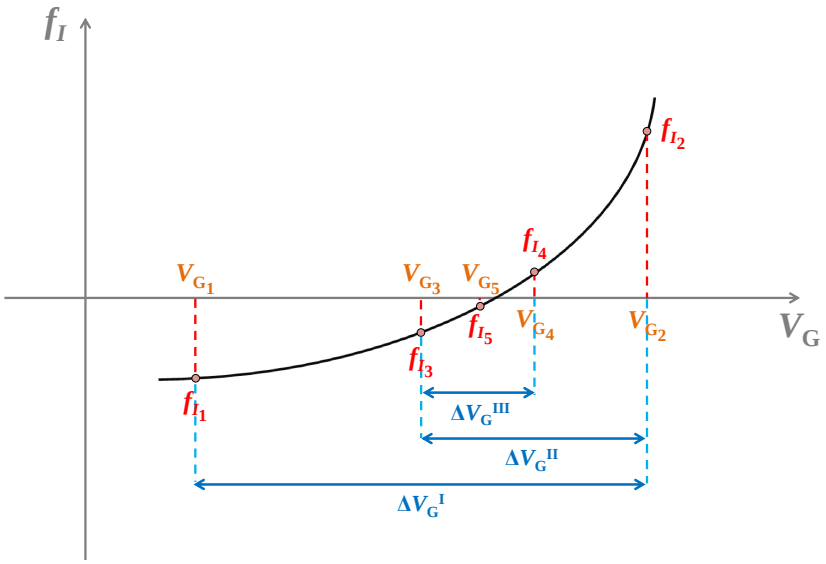

Figure 2.12: Bisection algorithm for the solution of Equation 2.71. The method is based on consecutive partitions of the V_{G} range. Every iteration implies at least one steady state electrical simulation for the evaluation of $I_{\mathrm{comp}}(V_{\mathrm{G}_i})$.

not easily invertible, such a problem is solved using a root-finding algorithm. The algorithm used in the simulator is the *bisection method* and it has been implemented by means of APDL in the electro-thermal simulation routine. Therefore, when simulating SP switch, the electrical iterations are not only necessary to take into account a possible de-biasing effect, but also to find the V_{G} value that satisfies the following equation:

$$|f_I(V_{\mathrm{G}})| = |I_{\mathrm{imp}} - I_{\mathrm{comp}}(V_{\mathrm{G}})| \leq I_{\mathrm{tol}} \tag{2.71}$$

where I_{tol} represents the maximum allowed discrepancy between I_{imp} and I_{comp}. This parameter is set before every simulation according to the required accuracy and eventually it can be chosen equal to the one defined in Equation 2.59.

The bisection method is based on consecutive geometrical division of the V_{G} search range until the V_{G} value which satisfies the Equation 2.71 is found (see Figure 2.12) [94]. For every V_{G} value one or two electrical simulations have to be run in order to determine the values of f_I at the considered gate volt-

age values, therefore also in this case several electrical iterations (simulations) have to be performed. An eventually more effective root-finding algorithm could be implemented in place of the relatively slow bisection method. Since the goal function f_I is always monotonic, a derivative-based method (such as the Newton–Raphson algorithm) or a secant method would easily find the solution of the nonlinear equation within a shorter computational time. The usage of those methods ensures a significantly faster convergence to the solution, therefore allowing a lower number of electrical simulations required by the root-finding procedure and, consequently, resulting in a less time demanding computational effort.

Once the V_{G} value has been found, the resistivity values associated to the channel elements can be computed by means of Equation 2.58 and the implemented algorithm continues with the electro-thermal simulation step according to the scheme reported in Figure 2.9.

With the described algorithm the value of V_{G} can be evaluated for every time step t_k. Eventually the values of V_{G_k} at every time step can be stored during the electro-thermal simulation run so that at the end of the computation the reconstructed V_{GS} waveform can be returned as a simulation output. Hence, the electro-thermal simulation scheme used to simulate SP devices requires as inputs the drain current and the drain-source voltage drop and returns as output the gate-source voltage, the potential and the temperature distributions, i.e. for every iteration step:

$$\left\{ \begin{array}{c} V_{\mathrm{DS}_k} \\ I_{\mathrm{D}_k} \end{array} \right. \quad \Rightarrow \quad \begin{array}{l} V_{\mathrm{GS}}(t_k) \\ \phi(\boldsymbol{x}, t_k) \\ T(\boldsymbol{x}, t_k) \end{array}$$

2.4 Simulator Validation

2.4.1 Test chips

In this work two main test structures have been used for:

- Having reasonable reference FE models of real power MOSFETs;
- Experimentally verifying the electro-thermal interaction in power MOSFETs;
- Validating the simulation results with experimental measurements.

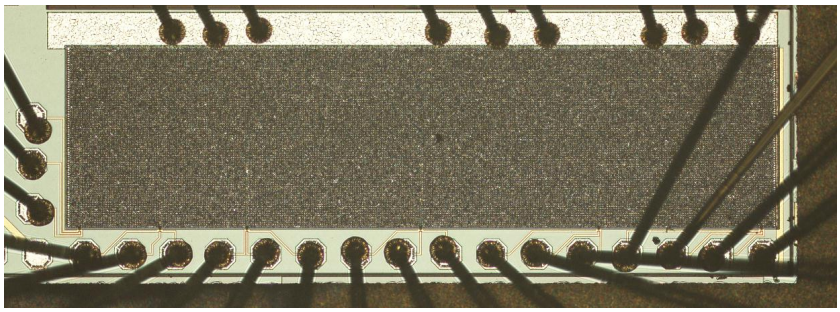

Figure 2.13: Active area of the DUT1. In this device 9 source wires are bonded in 3 rows of 3 wires each on an inactive region of the device placed at the top of the active area.

Both test devices are power MOSFET test chips realized using 6.4 µm cell pitch Trench technology (Paragraph 1.1.1). Details about the elementary cell of this technology have been presented by Kadow et al. in [70].

The first device (called *Device Under Test (DUT) 1*) is a rectangular power MOSFET featuring about 1 mm^2 active area, the area of the elementary cell is given by:

$$A_{\text{cell}} = {p_{\text{cell}}}^2 = (6.4\ \mu\text{m})^2 = 40.96\ \mu\text{m}^2 \tag{2.72}$$

Therefore the device includes roughly $24 \cdot 10^3$ elementary cells. In this MOSFET nine wires for the source contact have been bonded on an inactive region of the device located on the top of the active area, as it can be seen in the decapsulated sample shown in Figure 2.13. The other wires visible in the picture are used to contact temperature sensors integrated in the active area of the device, details on these sensors will be given in Paragraph 2.4.3. The source potential conducted by the wires reaches the epitaxial layer by means of a relatively thin copper (Cu) metallization and then, through the Inter-Layer Dielectric (ILD), by means of the *metal-1* contact (which is made of aluminum) and tungsten (W) plugs according to the *Damascene process* [95]. SEM cross-section pictures of the described metal stack are shown in the Figure 2.14.

The second device used in this thesis (*DUT2*) has been realized using the same Trench technology mentioned above with the same metal stack shown in Figure 2.14, but it is much bigger than the previous one. Since its active area is about 9 mm^2, this MOSFET includes approximately $220 \cdot 10^3$ elementary cells. A picture of a decapsulated sample is reported in Figure 2.15, there the reader should notice the source wires bonded on the active area and a temperature sensor placed in the middle region of the device.

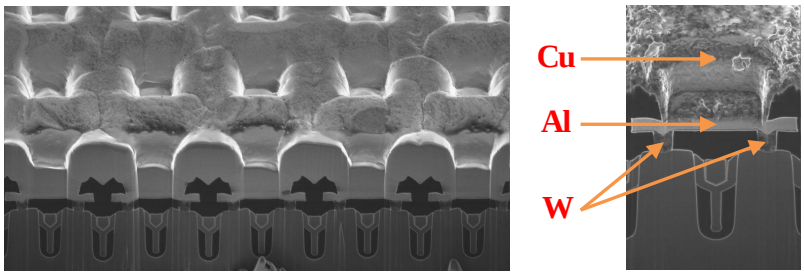

Figure 2.14: SEM picture of the metal stack structure. Figure on the right-hand side shows a detail of the ILD-interconnection stack where also the trench geometry within the epitaxial layer is visible. Darkest gray color represents oxides and BPSG.

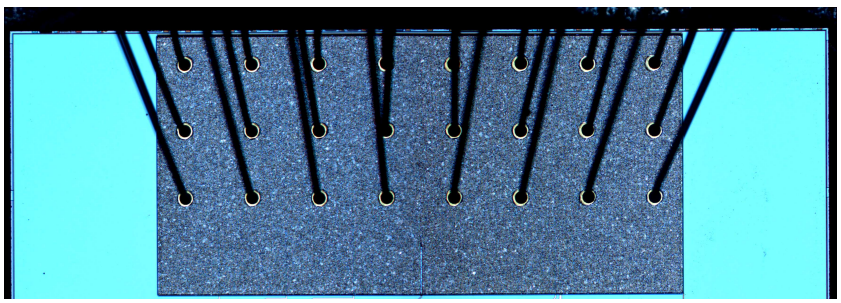

Figure 2.15: Active area of DUT2. In this device 24 source wires are soldered on the active area.

FE models of both test devices have been realized using APDL. Particular care has been used concerning the mesh strategy and the definition of the material properties of the elements:

- The 3-D models have been meshed using only hexahedron elements. The elements feature a quadratic (second order) basis function in order to achieve a higher solution accuracy at the node sites;
- Where possible, the active layer of the device has been meshed using a regular mesh, i.e. every element within the active area has square shape and it includes an integer number of elementary cells;
- The relatively thick silicon substrate has been meshed using an *exponential strategy*, namely the size of the elements along the vertical direction grows exponentially from the element layer nearby the epitaxial region to the substrate bottom, according to the following formula:

$$z_n = \mathbf{A}[e^{\mathbf{B}n} - 1] \tag{2.73}$$

  Where z_n is the vertical coordinate of the element n, $\mathbf{A}$ and $\mathbf{B}$ represent two coefficients which have been chosen according to the substrate thickness and to the desired size of the first element. Such a meshing strategy is particularly suitable because it allows a higher element resolution in the region where both the electrical and the thermal field gradients are larger, i.e. in the epitaxial layer, and a lower element resolution where the gradients are smaller, i.e. close to the bottom of the substrate. An example with 10 elements is reported in Figure 2.16;
- The epitaxial layer has been realized using a layer of *channel elements* which have been already described in Paragraph 2.3.1. Those elements are characterized by an anisotropic electrical resistivity so that in simulations they conduct the current only along the z (vertical) direction according to the Equation 2.58, while the electrical conductivity along *x* and *y* have been set to 0. This modelling approach represents a simplified design of the epitaxial layer, a more accurate one takes into account the thermal impact of the trench structures within the epitaxial layer and will be presented in the Chapter 3;
- The ILD layer shown in Figure 2.14 has been modeled using a layer of elements featuring equivalent material properties. This means that the electrical and the thermal material properties are equivalent to the

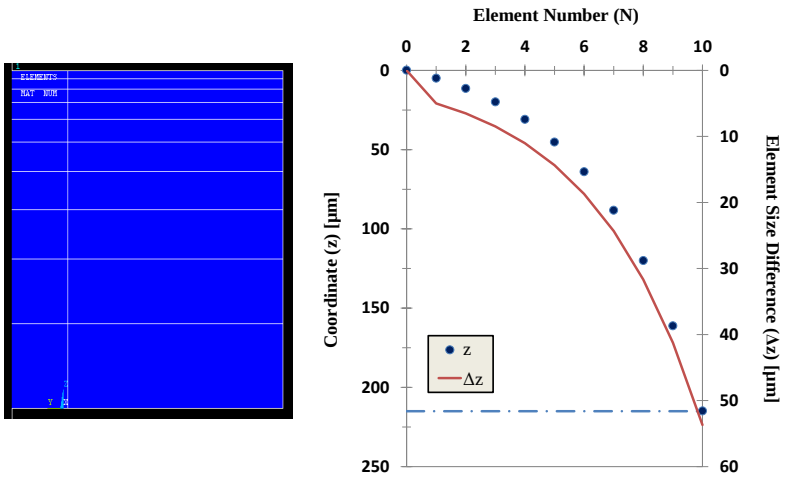

Figure 2.16: Element vertical size along the substrate thickness is defined according to an exponential function. In this example 10 elements are used to mesh the substrate of DUT1 and DUT2. Dots in the graph represent the element coordinates, the red curve describes the size-difference among contiguous elements.

ones that could be found if all geometrical details of the complex metal stack structure would be geometrically modeled in the FE model of the device [22]. An in-depth analysis of the methodology used for assessing equivalent material properties of highly anisotropic microstructures in the device geometry will be treated in Chapter 3;

- Nonlinear material properties have been defined for silicon and for ILD elements. This means that the substantial temperature dependencies of the silicon thermal conductivity, specific heat and electrical resistivity have been taken into account when performing electro-thermal simulations. The reader can find all the material properties defined for the simulations presented in this work in the Appendix A.

The FE model representing the device reported in Figure 2.13 features about $19 \cdot 10^3$ elements. It has been created using a regular mesh for the active area, where every element of the epitaxial layer includes 8×8 elementary cells of the device (see Figure 2.17).

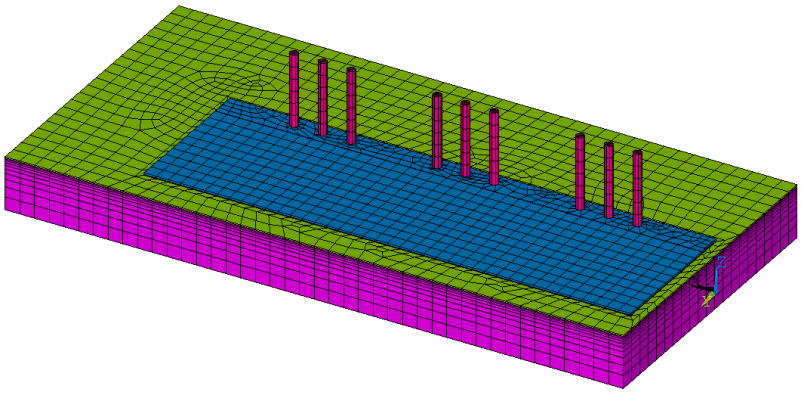

Figure 2.17: FE model of DUT1.

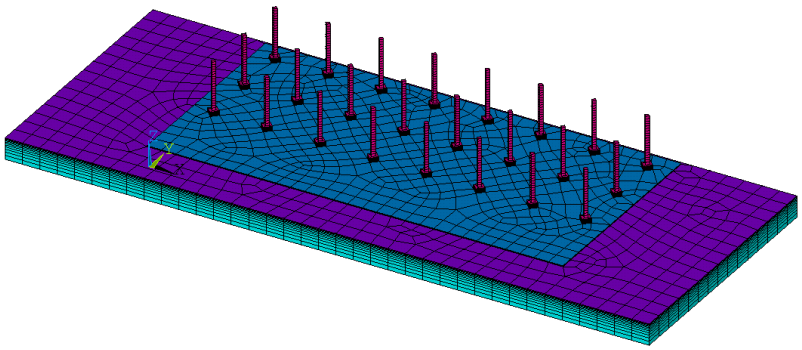

Figure 2.18: FE model of DUT2.

The FE model of the second device (Figure 2.15) does not feature a regular mesh for the epitaxial layer due to the bonding on active area. The presence of bond wires welded to the power metal upon the device active area does not allow the adoption of the regular mesh unless *contact elements* are used [22]. The whole model includes about $13 \cdot 10^3$ elements, the reader can find a picture of the meshed model in Figure 2.18.

2.4.2 Algorithm verification

Initially, electro-thermal simulations were performed in order to verify the correctness of implemented simulation concepts according to what is expected from the theory.

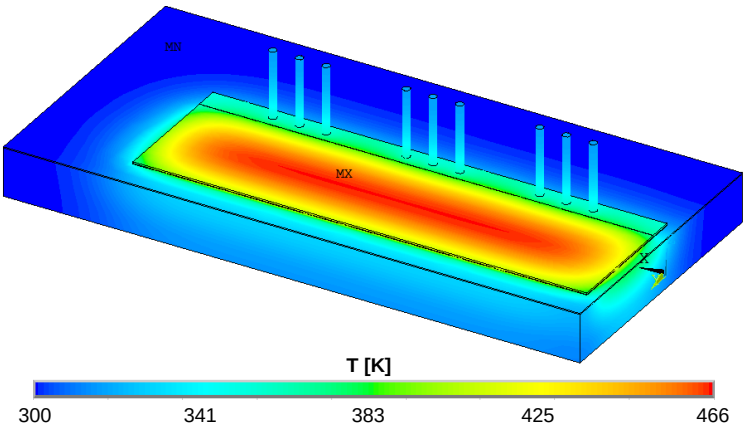

Figure 2.19: Temperature map of the DUT1 for the thermally unstable short circuit condition.

The FE model of the DUT1 has been used to simulate two kinds of short circuit conditions. A first transient simulation has been performed with the following settings:

- V_{GS} pulse with amplitude 2.2 V;
- V_{GS} pulse length of 500 µs;
- Constant V_{DS} of 20 V;
- Ambient temperature is 27°C.

The algorithm used to perform this simulation is the one described in Paragraph 2.3.1 which uses V_{GS} and V_{DS} waveforms as inputs. The operating condition defines a thermally unstable regime for the device operation, in fact the biasing V_{GS} amplitude is lower than V_{GS} at the TCP and this implies a negative α. In Figure 2.19 the surface temperature distribution of the test chip when the device reaches the highest temperature peak, i.e. at the end of the SC pulse, is shown. Figure 2.20 shows the temperature and current density distributions in the active area at the end of the pulse. It can be seen that the simulated operating condition leads to an inhomogeneous spatial distribution of both the temperature and the current density as expected from theoretical [12, 24] and experimental results [13, 96]. Furthermore, the hottest region of the active area

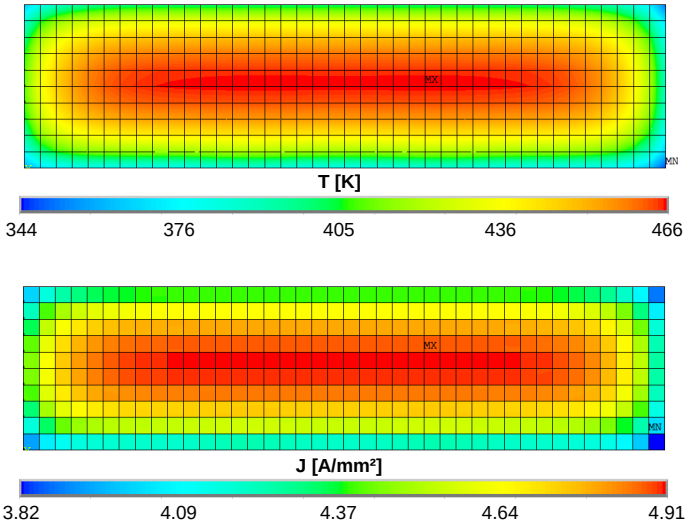

Figure 2.20: Temperature and current density maps in the active area of the DUT1 for the thermally unstable short circuit condition.

well matches with the region where the current density is the highest. This is a direct consequence of a positive α which leads to an increasing local current in the hotter region of the device. The temperature hot-spot emerges in the innermost region of the device due to the thermal coupling effect between the central cells. In fact, peripheral cells are surrounded by more efficient heat extraction paths, therefore generally they are colder than central ones.

At the end of the 500 µs pulse, the highest temperature in the device, reached in the innermost area of the epitaxial layer, is about 193°C. Due to the positive thermal feedback between temperature and current, a sufficiently longer pulse may lead the device to a thermal runaway failure (see waveforms in Figure 2.21). The change of the concavity in the temperature waveform can be used as a rough estimator for the triggering of the thermal runaway effect, in fact after the inflection point the hot-spot temperature starts to increase more rapidly [31]. The T_{HS} inflection point occurs after 660 µs at 233°C, at that time the current is 5.1 A.

In Figure 2.22 the reader can observe the emerging of the hot-spot associated with the crowding of the current in the central region of the device. After 1.62 ms the hot-spot within the device reaches roughly 763°C. At the same

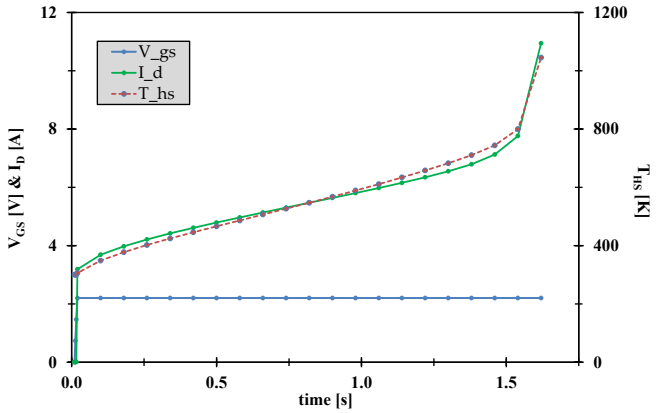

Figure 2.21: Simulated hot-spot temperature and drain current in DUT1 for the thermally unstable short circuit condition.

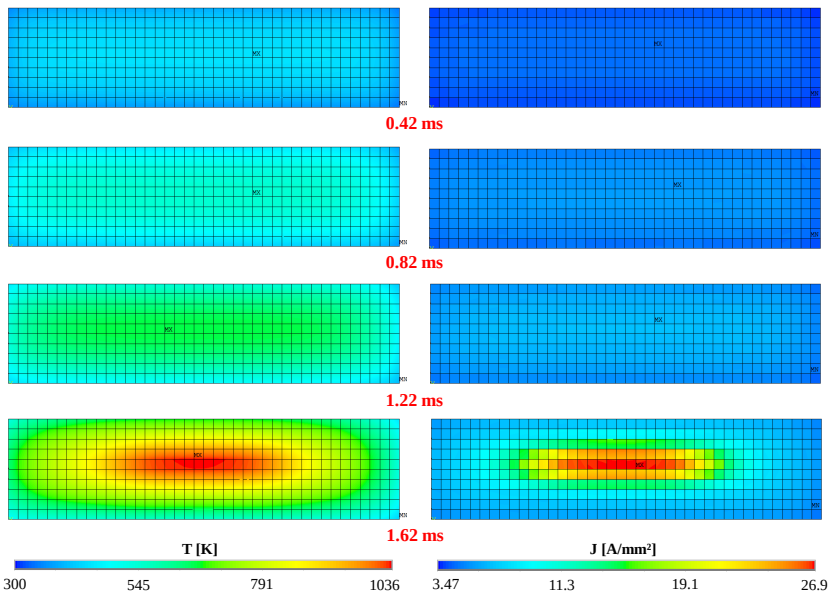

Figure 2.22: Temperature and current density maps in the active area of DUT1 at different time instants for the thermally unstable short circuit condition.

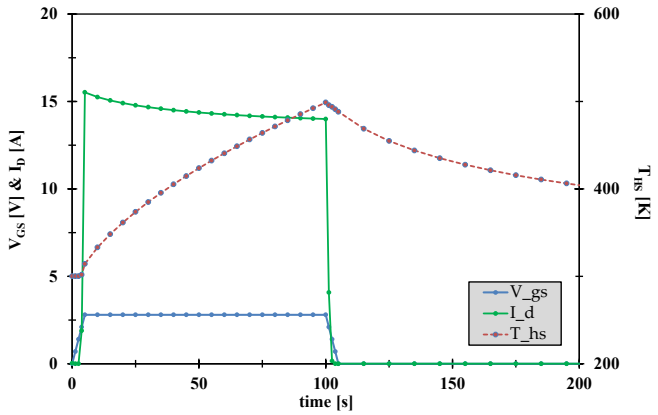

Figure 2.23: Simulated hot-spot temperature and drain current in DUT1 for the thermally stable short circuit condition.

time, the current in the innermost region is about 7 times bigger than the current in the peripheral regions, which means that about 11 A (see Figure 2.21) are flowing mainly within a small location of the whole active area available.

A second transient simulation has been performed with a thermally stable operating condition defined by:

- V_{GS} pulse with amplitude 2.8 V;
- V_{GS} pulse length of 100 µs;
- Constant V_{DS} of 20 V;
- Ambient temperature is 27°C.

In this case the V_{GS} value is greater than the one defined at the TCP, hence the device operates within the thermally stable regime, a proof can be found by examining the Figure 2.23 where the hot-spot temperature rising leads to a decreasing drain current. Furthermore, as expected, the negative value of α implies a more uniform temperature and current distribution in the device active area (see Figure 2.24). As seen in previous simulation results, also in this case the hotter elements are located in the innermost region of the device due to the thermal coupling among neighbouring elements. Nevertheless, colder elements conduct a slightly higher current density than hotter ones and this confirms the expected negative feedback between drain current and temperature and finally proves the correctness of the implemented simulation algorithm.

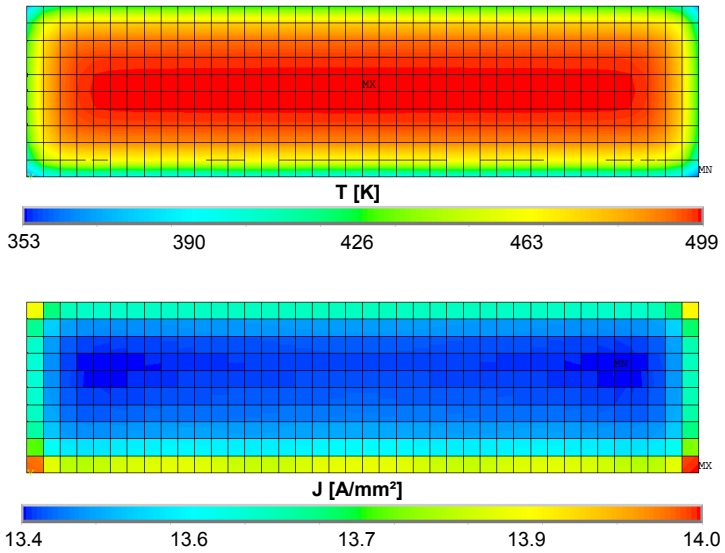

Figure 2.24: Temperature and current density maps in the active area of the DUT1 for the thermally stable short circuit condition.

Lastly, a comparison has been done by simulating two biasing conditions which define the same power dissipation. The first simulation has been performed using a current pulse of 5 A with the V_{DS} drop fixed at 40 V (*low-current/high-voltage* operation mode), as a consequence of the defined drain and voltage values 200 W are dissipated within the device. The second simulation defines the same power dissipation by means of a *high-current/low-voltage* operating condition, namely $I_{\mathrm{D}} = 20$ A and $V_{\mathrm{DS}} = 10$ V. Both operating conditions have been simulated using the algorithm presented in the Paragraph 2.3.3, which is suitable for cases where the current waveform is a known input (e. g. switching of an inductive load).

The low-current/high-voltage operation defines a thermally unstable operation of the device because in this pulse the drain current is always smaller than the current value defined at the TCP. On the contrary, the high-current/low-voltage operation represents a thermally stable condition. The results of a 400 µs pulse are depicted in the graph in Figure 2.25, there the rise of the hot-spot temperature in the unstable case is faster than the rise of the hot spot in the stable operation, even if the dissipated power in the considered electric pulses is equal. In particular, the hot spot temperature deviation increases

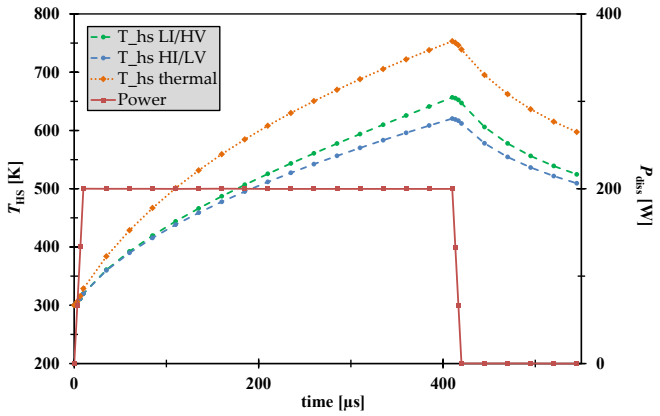

Figure 2.25: Electro-thermally simulated hot-spot temperature of low-current/high-voltage (green) and high-current/low-voltage (blue) operations and thermally simulated hot-spot temperature (orange).

in time and reaches already a peak of 36 K at the end of the pulse. In this graph the behavior of the orange curve represents the hot-spot temperature rising obtained with a pure transient thermal simulation. Such simulation consists of solving only the heat equation where the generation term Q_{V} in the Equation 2.6 is defined by the dissipated power within the active region of the device. Unlike electro-thermal simulation of power devices, performing a transient thermal simulation is an easy task since ANSYS software can solve such problem by FEM computing the corresponding FE formulation (described in Equation 2.38) without any further programming effort.

In the performed thermal simulation the temperature distribution in the device has been computed setting a uniformly distributed power, specifically 200 W, within the FEs composing the channel layer of the device model. The relevant temperature deviation between the thermally and electro-thermally simulated hot-spots underlines the importance of taking into account the interaction between thermal and electrical fields when simulating power device operation, the reader can find a further comparison and proof in [34]. Those results demonstrate that neglecting the electro-thermal interaction might lead to substantially wrong temperature estimations.

In Figure 2.26, the temperature distribution in the epitaxial layer at the end of the power pulse is depicted for both performed electro-thermal simulations. The same comparison at the end of the pulse is represented in Fig-

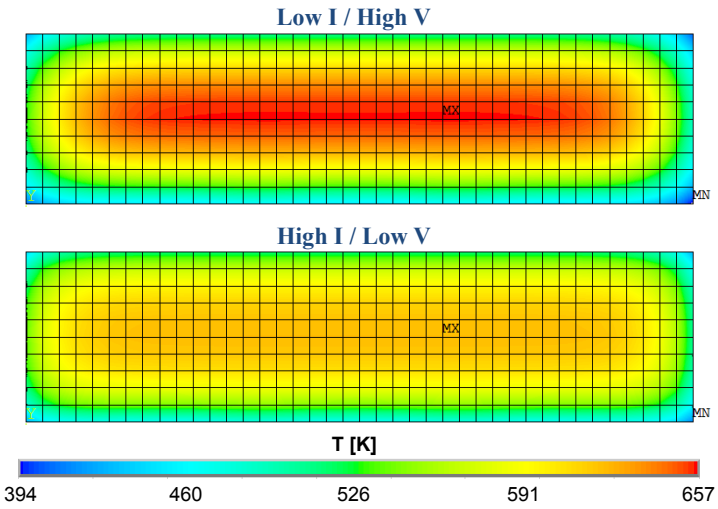

Figure 2.26: Temperature maps in the active area of the DUT1 for the low-current/high-voltage (top) and high-current/low-voltage operations (bottom).

ure 2.27 also for the current density. It is worth to notice that, as expected, the low-current/high-voltage operation defines quite uneven temperature and current distributions compared with the ones obtained by simulating the high-current/low-voltage operating condition.

Finally, further verifications and analyses of the developed simulation algorithm have been conducted by means of a simplified MOSFET structure featuring a single bond wire, the achieved results have been published in [35].

2.4.3 Integrated temperature sensors

The test devices presented in this thesis are equipped with temperature sensors integrated within the active area of the power device. In the DUT1, six temperature sensors are integrated in the MOSFET, their relative positions within the active region are depicted in Figure 2.28.

Every temperature sensor consists of a parasitic bipolar transistor implemented in the epitaxial layer of the device using a self-isolated common drain technology. The structure is designed by substituting a trench stripe with a $n+$ implantation and interrupting the source metallization to create the base and the emitter contacts of the bipolar sensor (see Figure 2.29). Therefore, every sensor has the base and the emitter terminals available for measurements while

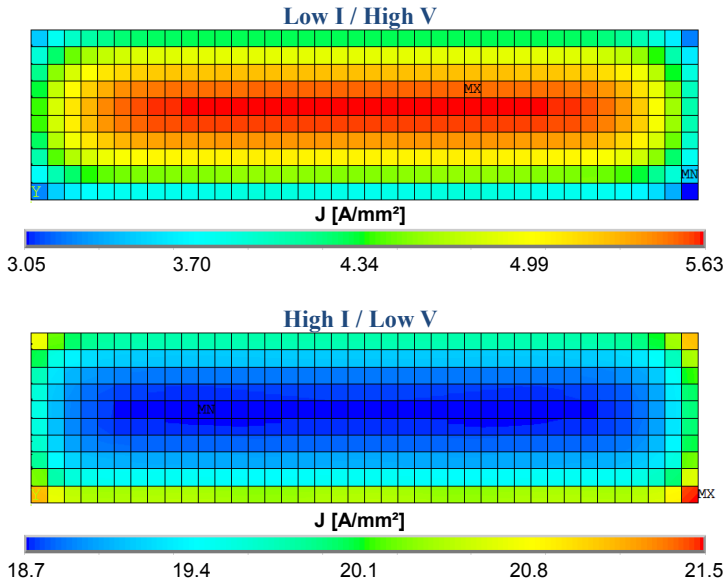

Figure 2.27: Current density maps in the active area of the DUT1 for the low-current/high-voltage (top) and high-current/low-voltage operations (bottom).

sharing its collector contact with the drain terminal of the power MOSFET.

Temperature sensors in the DUT1 have been calibrated using the circuit depicted in Figure 2.30. The calibration procedure consists of measuring the base-emitter voltage drop across the sensor at different ambient temperatures while a small calibration current is drawn out of the emitter contact. The Schottky diode placed across the base and the emitter of the sensor serves to protect the device from ESD failures.

The reader should notice that the sensor shares its collector terminal with the drain terminal of the MOSFET, hence during the normal power device operation the BJT corresponding to the sensor works inside the forward-active region of its output characteristics. Consequently, this kind of design results in a sensor which may be sensitive to the drain potential variation, the higher the V_{DD}, the bigger the leakage current crossing the collector towards the emitter. Therefore, the sensors have been calibrated for drain-voltage values used to bias the DUTs in the measurements presented in the Paragraph 2.4.4.

The characterization has been performed employing an airstream in a range from 10°C to 150°C obtaining a calibration curve for every sensor. This curve

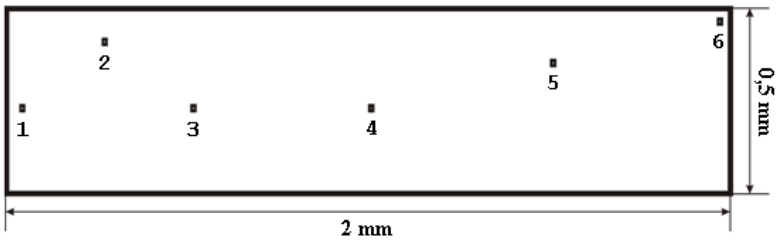

Figure 2.28: Temperature sensors positions within the active area of DUT1.

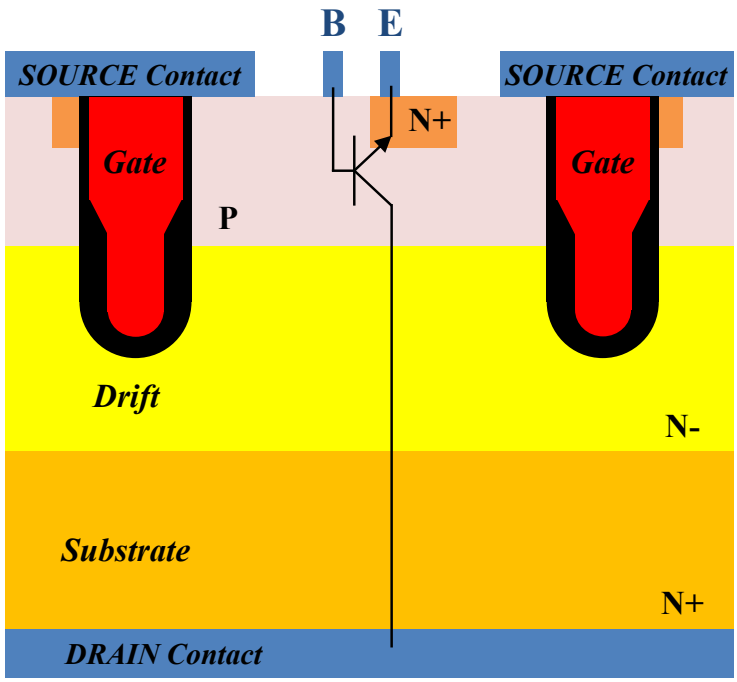

Figure 2.29: Bipolar temperature sensor integrated in the epitaxial region of DUT1.

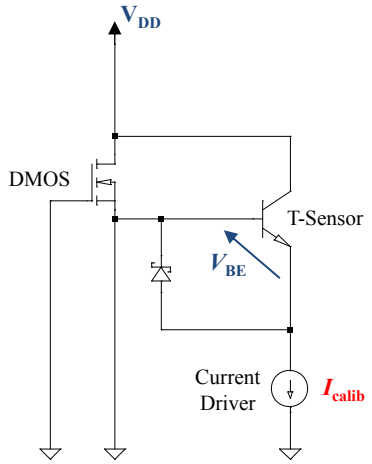

Figure 2.30: Schematic of the calibration circuit used for a temperature sensor integrated in the DUT1.

represents the temperature characteristic of the voltage drop across a *p-n* (base-emitter) diode and, generally, it shows a linear behavior up to 300°C [31, 43]. Hence the obtained experimental data have been fitted using the following linear model:

$$V_{\mathrm{BE}}(T_{\mathrm{s}}) = \gamma\, T_{\mathrm{s}} + \delta \tag{2.74}$$

Where γ typically has a value from -1 to -3 mV $^{\circ}\mathrm{C}^{-1}$, while δ represents a voltage offset (in mV) .

When calibrating such a sensor, the choice of a suitable value for the calibration current is extremely important. A higher current is generally recommended to contrast inaccuracies caused by increasing intrinsic carrier concentration at increasing temperatures, but this generally results in a higher self-heating of the sensor which may screen the real temperature value. The current value used here was 50 µA and it represents a good compromise between the two opposing effects within the considered temperature range.

As an example, Figure 2.31 reports the experimental data fitted according to Equation 2.74 for the sensor n. 4. The computed fitting parameters are: $\gamma = -1.5$ mV $^{\circ}\mathrm{C}^{-1}$ and $\delta = 838.7$ mV. Similar experimental data and fitting parameters have also been obtained for the other sensors integrated in the DUT1.

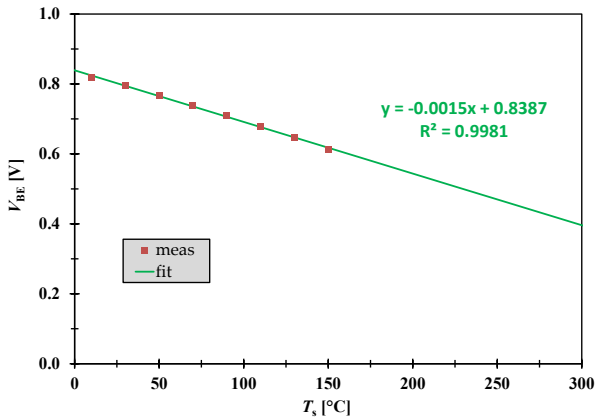

Figure 2.31: Experimental data and linear fitting for the sensor n. 4 integrated in the DUT1.

The DUT2 is also equipped with a bipolar temperature sensor. This sensor is placed in the lower part of the device active region, its position is visible in Figure 2.15. The design used for this sensor is more effective and robust compared with the one integrated within DUT1. In fact, here the sensor base terminal is internally connected to the drain of the power device, instead of being connected to the source of the MOSFET [43]. With such a circuit arrangement the sensor BJT is mounted in a diode configuration which allows a higher accuracy by reducing the leakage current of the base-collector junction. A similar design concept for reliably sensing the temperature in a power MOSFET has been also presented by Köck et al. in [14].

The temperature sensor in DUT2 has been calibrated according to the circuit depicted in Figure 2.32 in a temperature range from -40°C to 150°C using a Thermo-stream. Subsequently, the measured data have been fitted using the Equation 2.74 for allowing extrapolation of temperature values out of the used calibration range, both the experimental data and the linear fitting curve are reported in Figure 2.33.

2.4.4 Comparison of measurements versus simulations

A comparison between simulation results and experimental measurements is presented in this paragraph with the goal of validating the electro-thermal simulator accuracy. Several short-circuit measurements have been conducted us-

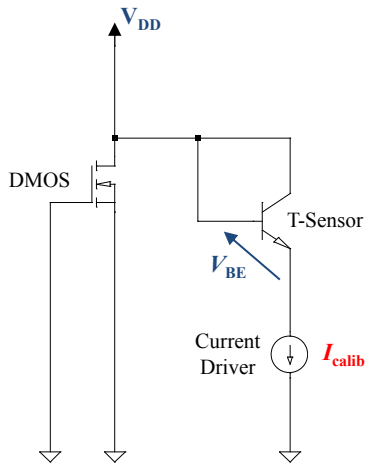

Figure 2.32: Schematic of the calibration circuit used for the temperature sensor integrated in the DUT2.

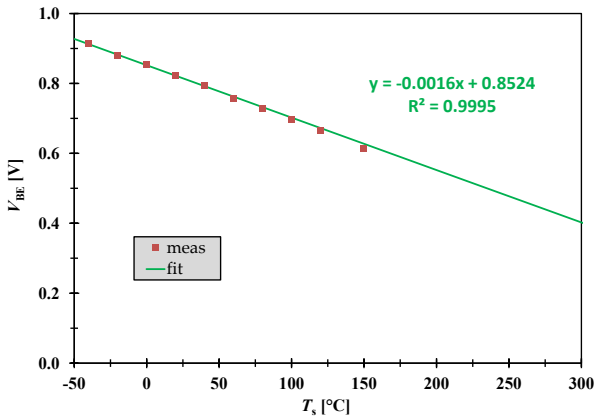

Figure 2.33: Experimental data and linear fitting of the sensor integrated in the DUT2.

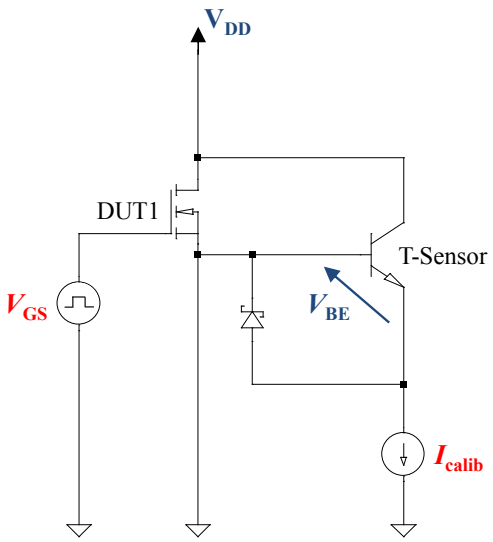

Figure 2.34: Schematic of the DUT1 measurement setup used for the experimental validation of the algorithm for power MOSFET simulations.

ing DUT1 and DUT2 with their integrated temperature sensors. The circuit scheme depicted in the Figure 2.34 has been used to test the DUT1 power MOSFET with constant V_{GS} and V_{DS} voltages and at the same time measure the base-emitter voltage drop across every integrated temperature sensor. During the measurement procedure an oscilloscope has been used to record and store the signal waveforms, an example is reported in Figure 2.35. Subsequently, the recorded V_{be} waveform samples have been converted to temperature values using the sensor calibration curves previously obtained (see Paragraph 2.4.3 and Figure 2.31). Being able to measure the V_{be} waveform for only one temperature sensor at a time, six pulses (one pulse for every sensor) of the same SC condition have been applied to the test device. We carefully applied the pulses sequence ensuring that after every stress pulse the device reached the thermal equilibrium with the ambient temperature before applying the next pulse. The short circuit condition defined for the DUT1 implies a thermally unstable regime and is specified by the settings below:

- V_{GS} pulse with amplitude 2.2 V;

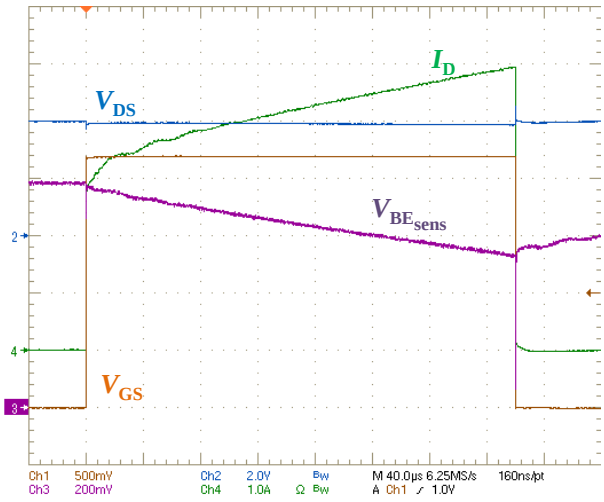

Figure 2.35: Example of oscilloscope waveforms detected during a short circuit pulse applied to the DUT1.

- V_{GS} pulse length of 300 µs;
- Constant V_{DS} of 40 V;
- Ambient temperature is 25°C.

Thereafter, with these settings an electro-thermal simulation was executed using the algorithm described in Paragraph 2.3.1 and the output results, namely the temperature values at the sensor locations and the device drain current, were compared with measurements. In the Figure 2.36 the simulated current well matches the measured one, here the biggest deviation is observed soon after the beginning of the pulse, i.e. at 20 µs, and is roughly 13%. This deviation is mainly due to setup parasitics not considered in the simulated model of the device. However, this represents a transient effect and does not have any influence on the thermal and electro-thermal behavior of the device. This is confirmed by comparing the time evolution of the measured and the simulated sensors temperatures, these comparisons are reported in the graphs in Figure 2.37. Moreover, in Figure 2.38 the temperature field in the active area at the end of the pulse is shown by highlighting the relative percentage of deviations between measured and simulated temperatures at the sensor locations. In order to get relative percentage errors independent from the used tempera-

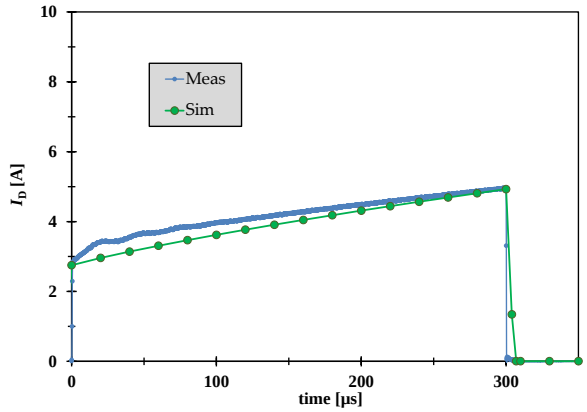

Figure 2.36: Comparison between measured (blue line) and simulated (green circular markers) drain current in the DUT1.

ture scale (kelvin and degrees Celsius), the relative errors have been computed using the deviation between the temperature value and the ambient temperature:

$$\Delta T_{\%} = \frac{|(T_{\text{sim}} - T_{\text{amb}}) - (T_{\text{meas}} - T_{\text{amb}})|}{T_{\text{meas}} - T_{\text{amb}}} \cdot 100 \tag{2.75}$$

In Figure 2.38 the reader can notice that the largest deviation is about 17% (in the sensor n. 2), while the smallest deviations, i.e. less than 1%, are located in the hottest region of the device (sensor n. 3, n. 4 and n. 5).

In the end, due to the good match achieved between measurements and simulations for a significant number of temperature sensors covering the whole active area of the employed test chip, we can conclude that the results accuracy of the developed algorithm has been successfully verified.

A final verification of the simulator proposed in this thesis has been conducted for validating the simulation algorithm developed for SP devices and described in Paragraph 2.3.3. Experimental measurements have been performed with two different versions of the DUT2 power MOSFET. The first version (called DUT2-A) has been fabricated employing a state-of-the-art technology currently used for commercial SP devices available on the market. The second version of device under test (DUT2-B) has the same chip size, active area dimensions, number of bond wires and geometries as DUT2-A, but it has been fabricated using an improved device technology which is characterized by a higher current capability, namely a higher g_{m}. FE models of the two DUT2

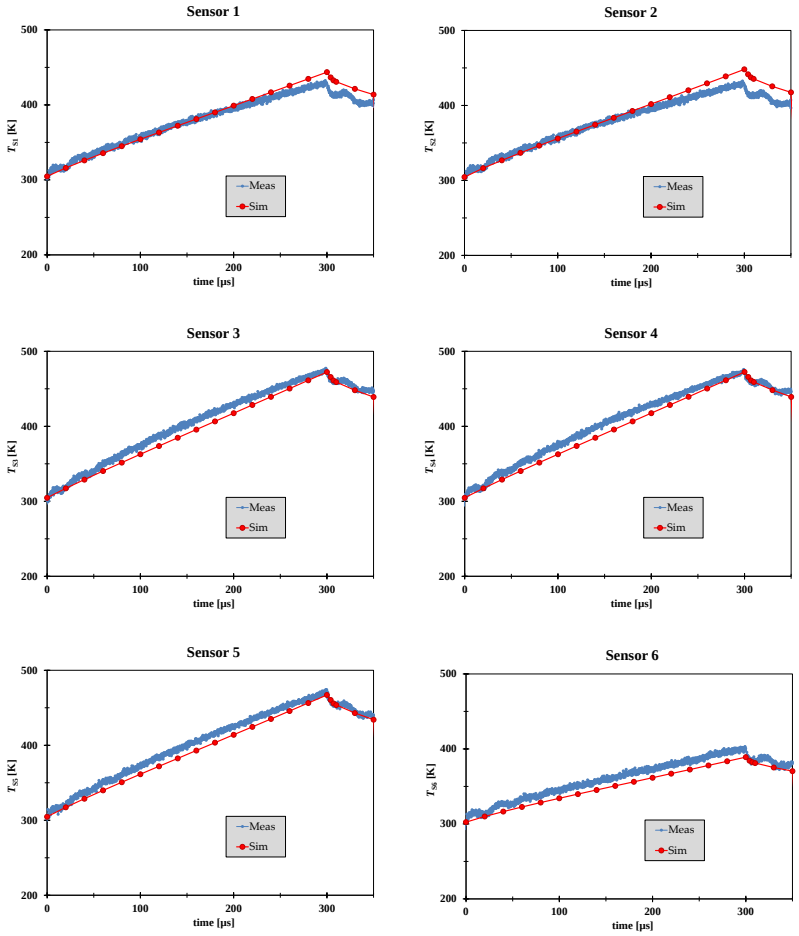

Figure 2.37: Comparison between measured (blue line) and simulated (red circular markers) sensors temperatures in the DUT1.

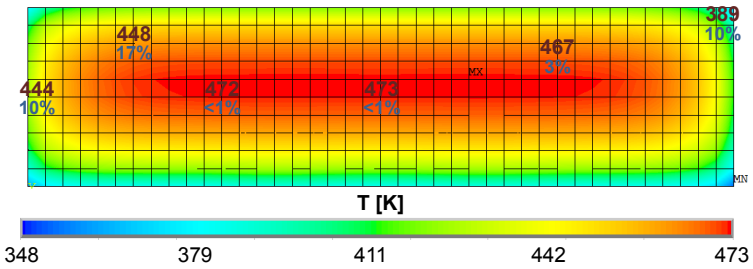

Figure 2.38: Simulated temperature distribution in the active region of DUT1 at the end of the short circuit pulse. Numbers indicate simulated temperatures (in K) and corresponding relative deviations with experimental values at the sensors locations.

versions have been used for a study that will be presented in the following of this thesis in the Paragraph 4.1.3. For now, in this paragraph the focus will be put only on the experimental validation of both designed devices models and the simulation algorithm.

Both test devices include in the middle of their active regions a bipolar temperature sensor (already described in Paragraph 2.4.3) visible in the Figure 2.15. In order to achieve accuracy when assessing temperature values by simulation at the sensor location, the FE models have been built by geometrically modeling the temperature sensor structure within the active area of the device.

The temperature sensor is physically realized within a small inactive area (i.e. the *finger* corresponding to the sensor) of the device active region. Along this finger the source power metal vias are missing and the metal-1 stripe is used only for contacting the sensor signal (i.e. the emitter voltage of the bipolar transistor) out of the active region. Therefore, in the FE model of the device both the ILD elements and the channel elements of the sensor finger have been modeled with an infinite electrical resistivity ρ_z along the vertical direction, while the vertical thermal conductivity λ_z in the sensor ILD has been set to zero. The FE model of the device including the temperature sensor finger is depicted in Figure 2.39.

The method employed here to model the temperature sensor within the FE description of the device is a simplified application of the homogenization technique, i.e. a multi-scale modeling approach suitable for accurately describe microstructures within the substantially bigger device geom-

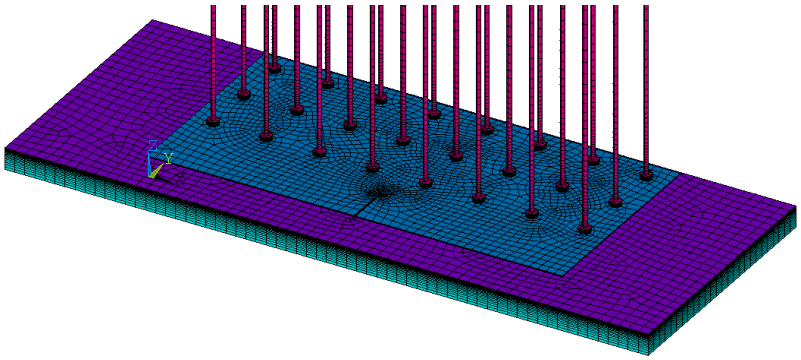

Figure 2.39: FE description including the model of the integrated temperature sensor of DUT2 (A and B).

etry [22, 97]. This method, compared with other available approaches such as non-matching grids, represents a good compromise between the modeling effort and resulting accuracy.

Measurements of currents and temperatures have been performed with both DUT2-A and DUT2-B using the test system proposed in [49, 50]. The experimental setup consists of a current driver purposely designed for testing SP switches with arbitrary current shapes. For the measurements presented below the driver has been used to inject into the drain terminal of the DMOS a triangular current pulse. Such pulse shape emulates the discharge of the magnetic energy accumulated inside a pure inductive load, such as valves or motors, and it represents a typical operating condition of SP devices working in the automotive electronic environment. In the schematic depicted in Figure 2.40 the gate potential is adjusted according to the instantaneous drain current value by means of a Zener-clamping net which includes a Zener diode and a resistor $\mathrm{R_G}$. The drain potential is defined by the Zener diode while the gate potential is defined by the voltage drop across the resistance $\mathrm{R_G}$. During the test pulse the base-emitter voltage drop of the integrated bipolar sensor has been measured by drawing out of the sensor emitter terminal the previously established calibration current. Finally, the temperature values have been extracted from the V_{BE} waveforms by means of calibration curves (see Figure 2.33).

Two distinct test pulses have been used to stress the devices at the ambient temperature of 125°C:

- $I_{\mathrm{D_{peak}}} = 10\mathrm{A}$, $V_{\mathrm{DS}} = 45\mathrm{V}$ and $T_{\mathrm{pulse}} = 500\mu\mathrm{s}$;
- $I_{\mathrm{D_{peak}}} = 40\mathrm{A}$, $V_{\mathrm{DS}} = 45\mathrm{V}$ and $T_{\mathrm{pulse}} = 500\mu\mathrm{s}$.

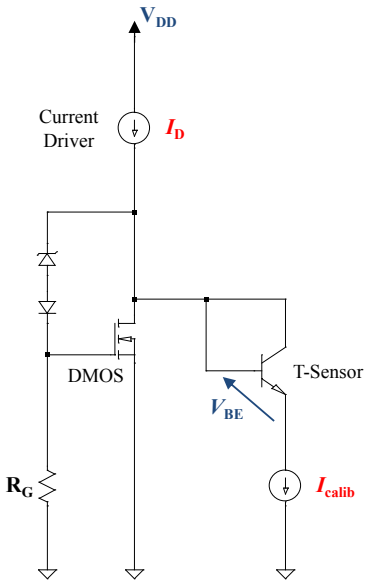

Figure 2.40: Schematic of the DUT2 (A and B) measurement setup used for the experimental validation of the algorithm for SP switch simulations.

In Figure 2.41 and Figure 2.42 comparisons between measured and simulated temperatures are reported for DUT2-A and DUT2-B, respectively. In these pictures it can be seen how the simulated time evolutions of the temperature in the sensor match with a reasonably good agreement the experimental ones for both devices and for the two test pulses, finally proving and validating the developed simulation algorithm.

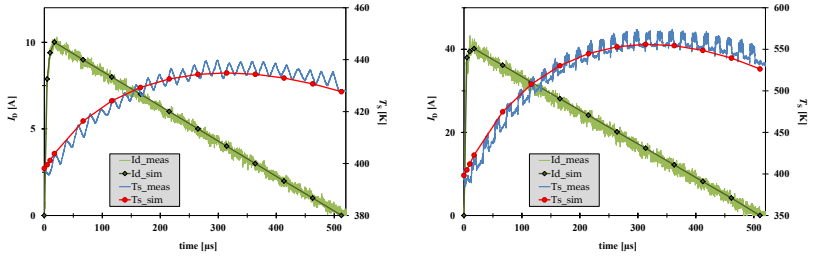

Figure 2.41: Comparison between measured (blue line) and simulated (red circular markers) sensor temperature in the DUT2-A for current peaks of 10 A (left-hand side) and 40 A (right-hand side). Light green curve describes the measured drain current while dark green rhombuses represent the simulated current.

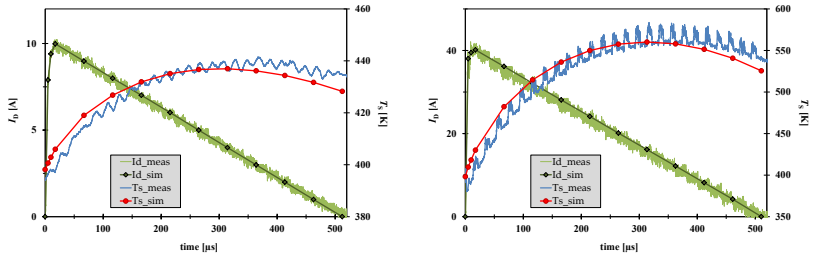

Figure 2.42: Comparison between measured (blue line) and simulated (red circular markers) sensor temperature in the DUT2-B for current peaks of 10 A (left-hand side) and 40 A (right-hand side). Light green curve describes the measured drain current while dark green rhombuses represent the simulated current.

Chapter 3

Device Modeling

Common power semiconductor devices consist of small entities, called elementary cells, which implement the physical working principle of the semiconductor device to which they belong. Elementary cells are connected in parallel within the device so that the total device current is the results of the summation of all current contributions due to every single elementary cell. For example, in modern vertical power MOSFETs every elementary cell shares with all other device cells the drain terminal (because they are realized on the same silicon substrate), the source terminal (due to the power metallization which covers the top of the device structure) and finally the gate terminal (implemented by polysilicon stripes which cross the device active region).

The technology treated in this thesis is state of the art for modern power MOSFETs [70]. It is characterized by 6.4 µm cell pitch, hence the active area associated to a single cell is 40.96 μm^2. Typically, the active region of a DMOS integrated within a modern SP switch is about few square millimeters. For instance, a DMOS featuring 2 mm^2 realized with the mentioned technology would include a few $50 \cdot 10^3$ elementary cells. Modeling a device using a geometry which describes all the device cells is normally not recommended because such a geometrical model would require an extremely large number of nodes/elements, limiting the device simulation due to hardware and/or software impediments related to the excessively long computational effort (and time) required. Therefore, when simulating the thermal/electro-thermal behavior of a power DMOS, the device geometrical model does not include all the geometrical micro-details present in the power device structure.

However, the core of the electro-thermal interaction, i.e. the dissipation of the electrical power and the electrical device response to temperature changes,

is mainly confined inside the epitaxial layer which is a small-pitched and often highly anisotropic area. Thus, in this case, the inclusion of geometrical details may be crucial.

In this chapter a method for modeling the detailed geometry of the repetitive pattern of elementary cells constituting the power device is presented. The adopted methodology exploits the homogenization approach, namely a method developed in computational science for representing a *multi-scale* FE model using simplified averaged structures. In this context the word "multi-scale" refers to a FE model characterized by geometrical structures which show substantially different feature sizes, e.g. the trench micro-structures or the metal interconnection stack between the power metal and the epitaxial layer (micro-scale) inside the geometry representing the whole power device (macro-scale). The usage of such an approach allows the FE modeling of semiconductor devices at a large scale, avoiding the FE description of their incorporated microstructures, thus employing a reasonable and acceptable number of elements/nodes. The FE model at large scale can be built using a conforming mesh and includes equivalent thermal and electrical behavior of the microstructures present within the device. This approach requires the prior creation of the detailed FE sub-model describing the micro-scale geometries in order to allow extraction of equivalent material properties by means of dedicated FE simulations.

In this chapter a FE model of the epitaxial layer of a trench power MOSFET is proposed [41]. The homogenization approach has been applied for estimating thermally equivalent material properties associated to the trench structures within the epitaxial layer. Afterwards, a modeling of the electrical material properties of the channel and of the drift region of the device is proposed. Finally, the impact of the proposed modeling on the device electro-thermal behavior is evaluated and compared to a commonly used isotropic model of the epitaxial region by means of electro-thermal simulations.

3.1 3-D Finite element model of a power MOSFET

3.1.1 Modelling the layers stack

In the thermal and electro-thermal simulations presented in this thesis a common approach has been used to model by FEs the device geometry describing a power MOSFET. The DMOS model has been built by stacking several elementary layers, where every layer represents a precise region of the device.

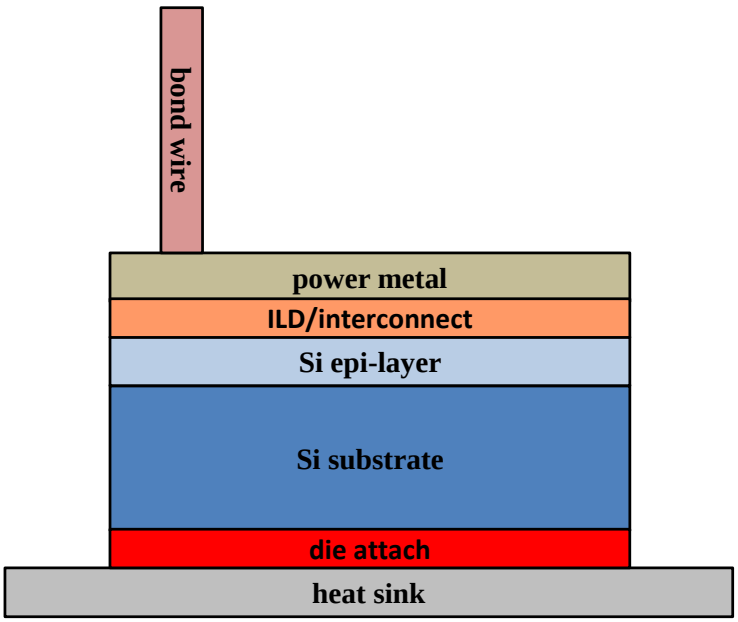

Figure 3.1: Layers stack of a generic power MOSFET for the FE geometrical description of the device.

Modern power switches are generally vertical devices, i.e. the current conduction path extends from the bottom to the top of the silicon substrate, therefore a power device can be modeled using the following basic layers starting from the bottom of the chip structure (see Figure 3.1):

- A generally thick layer (hundreds of micrometers) representing the heat sink (also called lead-frame) on which the silicon die is placed;
- A thin layer of few micrometers representing the die attach (glue, solder, diffusion soldering, etc.) used to fix the substrate to the heat sink;
- A generally thick layer of few hundreds of micrometers representing the silicon substrate;
- A thin layer representing the epitaxial region of device (tens of micrometers or less). This is the layer where the heat generation takes place due to the joule effect;

- A typically complex layer representing the interconnections structure together with the ILD. Such a layer is generally thin (tens of micrometers or less) and its thickness is proportional to the number of metal levels defined by the employed technology;

- A layer representing the power metallization of the device which is used to implement the source (or the emitter) contact of the power MOSFET (or of the power BJT/IGBT). The thickness of the power metal may vary from few micrometers up to tenths of micrometers;

- A layer made of tall cylinders representing the bond wires of the device. Bond wires can be located either directly on the active area of the device as in the DUT-2 (see Figure 2.13) or on its inactive area (see Figure 2.15).

Generally, when executing thermal or electro-thermal transient simulations featuring a small time scale (less than about 1 ms), the power device geometry can be created by avoiding both the lead-frame and the die attach layer assuming that the heat generated in the epitaxial layer does not spread down to the bottom of the substrate.

As mentioned already in the Paragraph 2.3.2, an analogous consideration is valid also for the geometrical modelling of the package. In fact, when performing short transient simulations, the package geometry is not necessary due to the very low thermal conductivity associated to the mould compound.

3.1.2 Temperature dependent material properties

In Paragraph 2.2.2 the FE formulation of the electro-thermal problem has been derived for the simplified case of linear material properties, namely the thermal conductivity tensor $\tilde{\lambda}$, the specific heat c, the density ρ and the electrical conductivity tensor $\tilde{\sigma}$ have been considered as temperature independent. When performing electro-thermal simulations of power MOSFETs in critical operating conditions (such as short circuit, overload, clamped/un-clamped discharge of an inductive load, etc.) the temperature rise within the device may easily cover hundreds of Kelvin even in few microseconds. For such a wide temperature range the involved thermal and electrical material properties may vary substantially. For example, over a temperature range from 25°C to 300°C, the thermal conductivity of the silicon reduces by about 40%, while its substrate electrical resistivity icreases by about 30% (for *n*-doped silicon with a doping

level of $2 \cdot 10^{19}\text{cm}^{-3}$). Therefore, temperature dependencies of material properties used in the device FE description have to be taken into account in order to achieve an adequate level of accuracy in simulation results.

For all the simulations presented in this work, non-linear material properties parameters have been used by specifying a table reporting the absolute value of the temperature with the respective material property value, for the following parameters:

- Thermal conductivity, specific heat and electrical resistivity of the silicon in the substrate and in the epitaxial layer [98];
- Thermal conductivity and specific heat of the gate polysilicon [99];
- Thermal conductivity of the field and gate silicon oxides (used in the trench);
- Thermal conductivity of the BPSG (used in the ILD region).

Values of employed thermal material properties have been reported in Appendix A.

3.2 Modeling of the epitaxial layer in the trench technology

3.2.1 Modeling anisotropic microstructures: homogenization

Homogenization represents one of the sub-modeling approaches developed in computational science suitable for overcoming computational limitations introduced by multi-scale problems [37]. Originally, this method has been developed for the analysis of porous and composite materials [38, 39, 40]. The goal of this approach is to determine equivalent material properties of complex heterogeneous micro-structures included in the macro-scale model. In particular, in this thesis the method has been used to assess thermally equivalent material properties of the epitaxial region in a trench technology.

The micro-scale model is described within the macro-scale geometry using equivalent layers which are built using a conforming mesh compatible with the mesh size used for the macro-scale geometry. In a thermal analysis, the aim of the homogenization method is to consistently describe the thermal properties relations between the micro and the macro-scale geometry. This means that the temperature field in the macro-model (i.e. the computational domain) must be

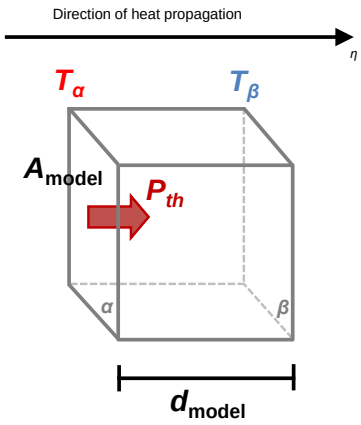

Figure 3.2: Scheme for the evaluation of the equivalent thermal conductivity of a generic model along an arbitrary direction (η) of heat propagation.

determined by taking into account thermal effects associated to the the micro-scale geometry if the latter would be described in detail within the considered computational domain. Effectively, the used homogenization scheme consist of:

1. Creating the FE model of the micro-scale structure using a proper (fine enough) mesh and defining its material properties;

2. Performing FEM simulations aimed at the evaluation of the equivalent material properties. Generally, those material properties are orthotropic, this means that if a Cartesian coordinate system is used, the evaluation should be done along the three axial spatial dimensions. From an operative point of view, we need to solve by FEM the heat equation in the steady state regime:

$$-\nabla\tilde{\lambda}\nabla T(\boldsymbol{x}) = Q_{\mathrm{V}} \tag{3.1}$$

   In this case, the equivalent material properties to be evaluated by homogenization are the thermal conductivity of the micro-model along x, y and z directions. Along the generic η direction, the way to proceed is to apply a Neumann BC (generation term) to one of the two model surfaces orthogonal to the η direction:

$$\boldsymbol{q}_{\mathbf{n}}|_{\alpha} = \overline{\boldsymbol{q}} \tag{3.2}$$

while on the opposite surface, a Dirichlet BC defining the ambient temperature (sink term) has to be applied:

$$T|_{\beta} = T_{\text{ambient}} \tag{3.3}$$

as schematically depicted in Figure 3.2. Subsequently, a thermostatic FEM simulation can be executed for determining the temperature distribution within the micro-model. If the material properties defining the FE micro-model are non-linear, several simulations should be performed for different ambient temperatures in order to take into account the thermal behavior of the micro-model at different temperatures.

3. Determining the equivalent material properties. In the thermal case, once all micro-model simulations have been performed for all directions, and in case of non-linear material properties for all temperatures, the equivalent thermal conductivity λ_{eq_η} associated to the generic direction η can be assessed by means of the following expression:

$$\lambda_{\text{eq}_\eta} = \frac{d_{\text{model}}}{A_{\text{model}}} \cdot \frac{P_{\text{th}_\alpha}}{T_\alpha - T_\beta} \tag{3.4}$$

where T_α and T_β represent the average temperature on the corresponding surfaces α and β orthogonal to η, i.e. the heat propagation direction. A_{model} and d_{model} are the area in m^2 and the model thickness along η in m, respectively. Finally, P_{th_α} is the power in Watt applied on the surface α. This term can be also expressed as a function of the associated heat flux q_α:

$$P_{\text{th}_\alpha} = q_\alpha \, A_{\text{model}} \tag{3.5}$$

The Equation 3.4 descends directly from Fourier's law, which describes the one-dimensional heat flux q as a function of the linear temperature drop along the generic direction of propagation η across an infinite and homogeneous (i.e. isotropic) rod:

$$-\lambda_\eta \frac{dT}{d\eta} = -\lambda_\eta \frac{\Delta T}{\Delta \eta} = q_\eta \tag{3.6}$$

Reworking Equation 3.6 together with Equation 3.5 leads to the definition of λ_{eq_η} reported in Equation 3.4. In this way a tensor $\tilde{\lambda}_{\text{eq}}$ can be determined for every ambient temperature with the goal of defining the

equivalent non-linear orthotropic thermal conductivity associated to the micro-model:

$$\tilde{\lambda_{\mathrm{eq}}}(T) = \begin{bmatrix} \lambda_{\mathrm{eq}_x}(T) & 0 & 0 \\ 0 & \lambda_{\mathrm{eq}_y}(T) & 0 \\ 0 & 0 & \lambda_{\mathrm{eq}_z}(T) \end{bmatrix} \tag{3.7}$$

4. Defining equivalent material properties in the macro-model. The obtained $\tilde{\lambda_{\mathrm{eq}}}(T)$ must be now employed to define the material property values of the elements (generated with the macro-model conforming mesh) representing the layer describing the micro-structures in the FE macro-model.

The homogenization approach represents one of the methods available for dealing with multi-scale problems. Its main drawback is given by the fact that micro-structures are replaced in the macro-model by a roughly meshed region according to the element mesh size used for the macro-model. In this sense, micro-model details become hidden inside the macro-model mesh, namely the approach does not allow to retrieve values of degrees of freedom within the micro-model when simulating the whole macro-model. In power devices, temperature peaks and gradients associated with thermal expansion are driving sources of most degradation mechanisms [5, 6, 7, 8, 9, 10, 11]. Therefore, the usage of the homogenization approach must be limited to cases where problem unknowns have to be determined in the macro-model keeping in mind that unknowns values within equivalent layers do not correctly describe the real field within micro-structures, but they are only useful for correctly determining their impact on the whole macro-model.

Another effective, but more complex, approach which overcomes this drawback is represented by the usage of non-matching grids (also called *mortar FEM*) in the device mesh. This method consists of meshing the macro and micro-model together by using two substantially different mesh sizes. In this sense, the connection between the macro-elements and micro-elements is realized using a modified FE formulation in which some elements can share one border with more element borders belonging to adjacent elements. The main drawback associated to this method concerns practical aspects connected to the effective implementation in the simulator, such as the handling of intersection operations at the non-matching interfaces or the definition of the discrete coupling operators. Furthermore, the mortar FEM theory is still a widely studied topic in the numerical computation community and it is not considered completely stable yet for all engineering problems. The reader may find valid

references for the analysis of numerical problems (e.g. simulations of wave propagation, thermal problems and coupled mechanical-acoustic problems) by means of non-matching grids in [82, 22, 100, 101]. In this work the mortar FEM approach has not been adopted, preferring to face sub-modeling problems in multi-scale geometries by means of the homogenization method only. We will see in the following paragraphs that the homogenization represents a valid approach when investigating by FEM simulations the electro-thermal behavior of power semiconductor devices.

3.2.2 Epitaxial layer stack

The homogenization procedure explained above will now be applied to determine the thermally equivalent material properties associated to trench structures implemented in the epitaxial layer of a power MOSFET.

When performing thermal or electro-thermal simulations of power devices the epitaxial region is generally represented with a single layer made of silicon in the device stack structures presented in Paragraph 3.1.1. Such a kind of modeling obviously neglects:

- The eventual orthotropic thermal effect of possible trenches structures;
- The electrical effects of the substantially different doping levels and doping gradients used to realize the device channel and an effective drift region according to the breakdown capability and to the R_{ON} requirement.

Based on the geometrical structure of the trench, the epitaxial layer has been decomposed in three distinct sub-regions. A picture of the trench obtained with a Scanning Electron Microscope (SEM) highlighting the sub-regions is reported in Figure 3.3. In this technology the trench is 4 μm deep, while the pitch of the elementary cells is 6.4 μm where every elementary cell includes 2 trenches. In the Figure 3.3 the silicon and the polysilicon appear in light gray, while the oxide regions, i.e. silicon oxide and BPSG, appear in dark gray. In the proposed geometrical modeling the thin oxide layer which separates the outer polysilicon (which is biased at the gate potential V_{G}) from the inner polysilicon (also called *filler*), has been neglected due to its very small thickness. The three sub-regions proposed can be distinguished as:

- The *channel region*. In this region the switching operating principle takes place. An electron channel is formed due to the creation of the inversion layer by positive gate biasing. This region is made of: 1) highly

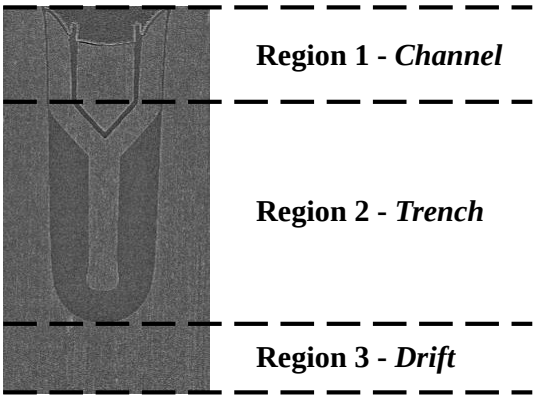

Figure 3.3: SEM picture of the trench within the epitaxial layer and definition of the sub-regions.

doped polysilicon used for both the trench filler and for the channel gate; 2) silicon dioxide, i.e the so called *gate oxide* and the field oxide; 3) BPSG, namely a particular kind of doped silicon oxide employed in the ILD to separate the metal-1 contact from the polysilicon gate contact. Such a geometry is clearly strongly anisotropic regarding the thermal and the electrical propagation.

- The *trench region*. In this region the channel ends and the well confined channel current starts to spread towards the drift region because the thin gate oxide progressively degenerates towards the thicker field oxide so that the gate biasing effect is substantially attenuated. An example of the current spreading within the trench region is shown in Figure 3.4, which has been obtained with a TCAD simulation. The so called trench region is made of doped silicon, doped polysilicon and silicon oxide and, as in the channel region, a strong thermal and an electrical anisotropy are present due to the field plate trench.

- The *drift region*. This is a relatively low doped silicon region. Here the current flows towards the silicon substrate. This region is made of silicon only and it does not show any kind of thermal and electrical anisotropy, though the current still flows mainly along the vertical direction.

In order to assess the thermally equivalent material properties associated

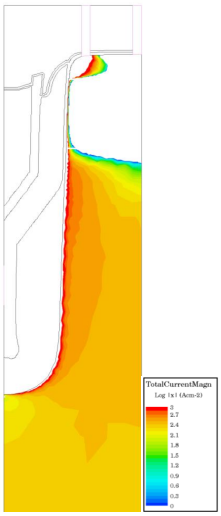

Figure 3.4: Current spreading inside the *mesa* region resulting from a two-dimensional TCAD simulation (courtesy of Stefan Decker).

to the epitaxial layer, 3-D FE models have been created in ANSYS for the detailed micro-models of the channel (depicted in Figure 3.5-A) and the trench region (depicted in Figure 3.5-B). The drift layer does not require any FE modeling since no anisotropies are present in this region.

3.2.3 Equivalent thermal material properties: thermal conductivity

The evaluation of the equivalent thermal conductivity associated to the channel and trench region has been performed by applying the homogenization approach. As explained already in the Paragraph 3.2.1, this method requires several FEM steady state thermal simulations for solving the static heat conduction equation along the three spatial directions.

Given the FE model of the structure under investigation, a superficial heat generation is imposed on one of the two model surfaces orthogonal to the direction of the heat propagation we want to investigate (for example, surface α in Figure 3.2). On the second surface β, a Dirichlet boundary condition defines the value of the ambient temperature. Finally, all the remaining surfaces of the model (all of them parallel to the direction of the heat propagation) are set

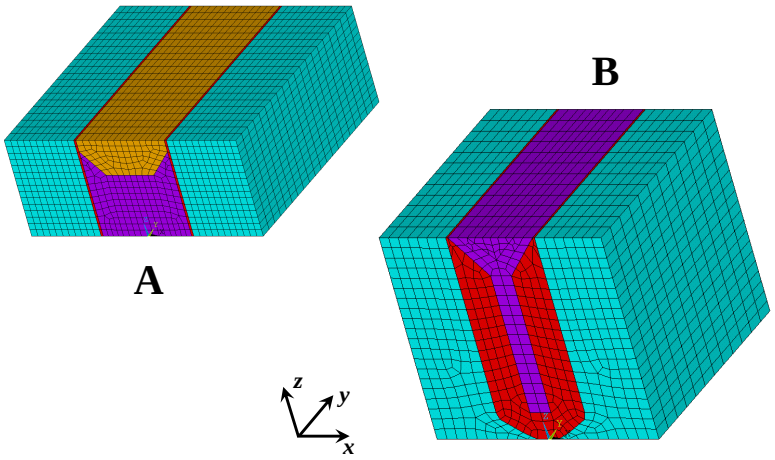

Figure 3.5: FEM models of both channel (A) and trench regions (B). Materials (colors): silicon (turquoise); polysilicon (violet); silicon dioxide (red); BPSG (orange).

to be adiabatic. For instance, to investigate the heat propagation along the *x*-direction in the channel region a superficial heat generation has been imposed on the right side *yz*-surface of the FEM model in the Figure 3.5-A while the ambient temperature has been fixed on the left side *yz*-surface.

After performing the thermal simulation, the temperature distribution in the whole structure is known. Hence, the equivalent thermal conductivity of the model in a certain direction can be directly determined according to Equation 3.4.

Since material properties of both detailed models are non-linear, the evaluations have been carried out using several T_{ambient} chosen in a wide range of temperatures relevant for applications of interest, e.g. short circuit pulses in thermally stable and unstable regimes. When performing thermal simulations for estimating non-linear equivalent thermal conductivity values, the choice of a small enough power value for the generation surface is of paramount importance. In fact, a big temperature rise induced in the detailed model may be affected by intrinsic non-linearities associated to component materials (e.g.

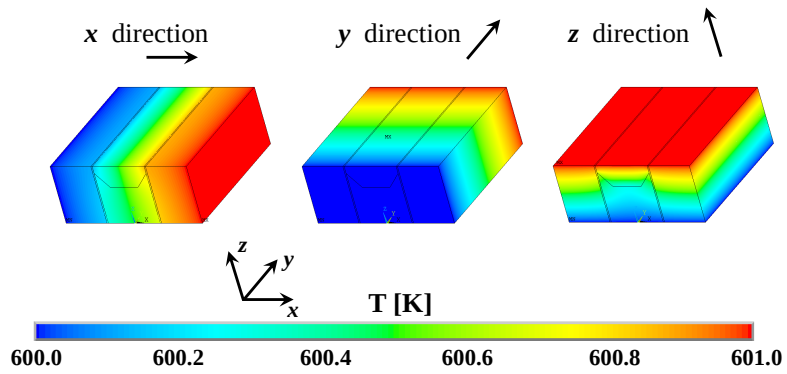

Figure 3.6: Temperature distributions in the channel model resulting from steady state thermal simulations along the three axial spatial directions at 300 K.

silicon temperature dependence). Therefore, a small power value Q_β for the generation surface has been set in order to obtain a ΔT of 1 K for all chosen ambient temperatures. Indeed, for such a small temperature gradient, thermal conductivities related to component materials do not vary substantially with the ambient temperature. For instance, the reader may observe in Appendix A that the λ temperature dependence of silicon within 1 K is absolutely negligible.

Finally the homogenization method has been applied for thermal characterization of both the channel and the trench regions. As example, Figure 3.6 shows the temperature distributions within the channel layer at ambient temperature of 600 K along the three spatial directions. Another example is depicted in Figure 3.7 for the trench layer at 300 K.

In both examples it is particularly interesting to notice the strong isolation effect introduced by the oxide trench regarding the heat propagation along the x direction. On the contrary, the thermal conduction along the y and z direction is only slightly diminished when compared to the λ of silicon. This effect can be seen in Figure 3.8, where the thermal characterization results, normalized to the value of the silicon thermal conductivity at 300 K, are graphically depicted. The reader should also notice that the equivalent estimated thermal conductivity along the y direction is slightly larger than the one along the z direction. This means that the most efficient way to extract the heat from the epitaxial layer does not correspond to the current flowing direction, but it is actually orthogonal to the latter.

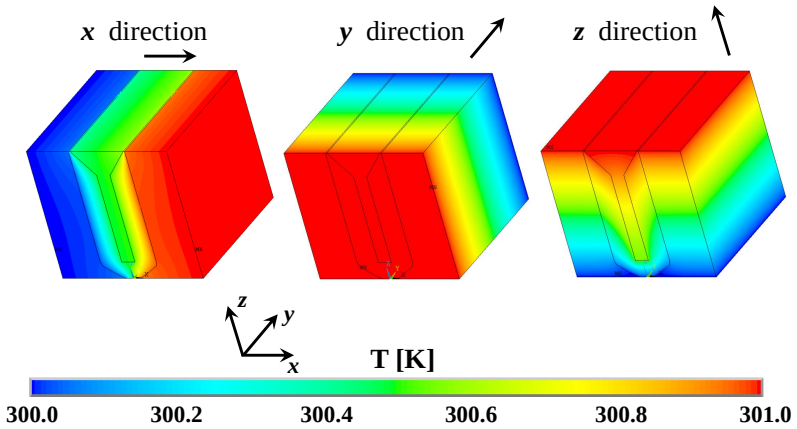

Figure 3.7: Temperature distributions in the trench model resulting from steady state thermal simulations along the three spatial directions at 300 K.

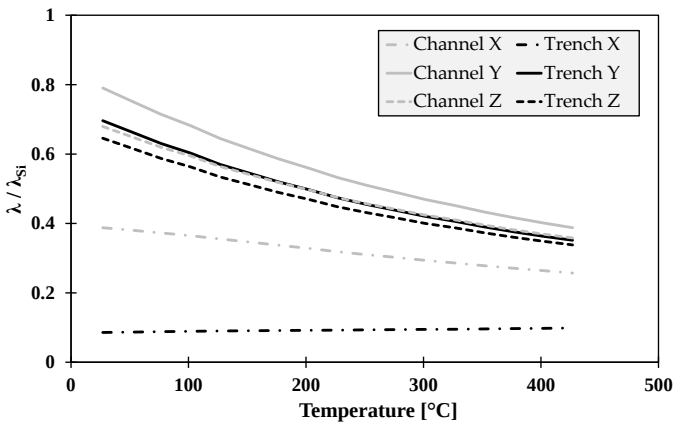

Figure 3.8: Equivalent orthotropic thermal conductivities as a function of the ambient temperature for the channel and the trench model. λ values are normalized to the value of the silicon thermal conductivity at 300 K.

3.2.4 Equivalent thermal material properties: thermal capacity

The assessment of thermally equivalent material properties of the channel and the trench region implies the evaluation of associated equivalent thermal capacities, since the developed electro-thermal algorithm solves by FEM the non-stationary heat equation (Equation 2.6).

The thermal capacity, or heat capacity, C_{th} is defined as the amount of heat required to change the temperature of a material by one Kelvin, thus C_{th} is normally expressed as:

$$C_{\text{th}} = \frac{\partial Q}{\partial T} \tag{3.8}$$

where Q is the amount of heat in Joules and T the absolute temperature in kelvin. Heat capacity is an extensive property, i.e. a physical property that scales with the size of the system. Commonly, it is more convenient to deal with the respective intensive property of C_{th} which is defined as the specific heat (also called specific heat capacity) c, namely a heat capacity per unit of mass m:

$$C_{\text{th}} = cm = c\rho V \tag{3.9}$$

where ρ is the density of the material and V its volume in m^3.

When solving transient thermal problems by FEM commercial software suites, in addition to the thermal conductivity the user has to assign values of ρ and c to every model material. When dealing with nonlinear thermal problems, these parameters should be eventually specified for different temperatures. Therefore, when considering a generic FE model made of two or more materials, the problem of the evaluation of an equivalent density ρ_{eq} and an equivalent specific heat c_{eq} arises. For such evaluations no physics-based definitions or formulations exist, hence an heuristic approach has been adopted.

The equivalent density has been estimated by applying the definition of the density to the whole FE composite model (system), namely the equivalent density is given by the total mass of the system divided by its total volume:

$$\rho_{\text{eq}} = \frac{\sum_{i=1}^{M} \rho_i V_i}{\sum_{i=1}^{M} V_i} \tag{3.10}$$

where M is the number of materials included in the considered system.

Similarly, the equivalent specific heat has been computed by evaluating the mass-weighted sum of the all specific heat contributions associated to every

material of the system:

$$c_{\mathrm{eq}} = \frac{\sum_{i=1}^{M} c_i \left(\rho_i V_i\right)}{\sum_{i=1}^{M} \rho_i V_i} \tag{3.11}$$

According to Equation 3.9, the proposed estimation of the equivalent specific heat can be interpreted also in another way, namely c_{eq} is given by the summation of all thermal capacity contributions $c_i m_i$ divided by the total mass of the system:

$$c_{\mathrm{eq}} = \frac{\sum_{i=1}^{M} C_{\mathrm{th}_i}}{\sum_{i=1}^{M} m_i} \tag{3.12}$$

Estimations given in Equation 3.10 and Equation 3.11 have been used to compute equivalent densities and equivalent specific heats respectively of the channel and trench model. It should be noticed that, contrary to the case of previously estimated equivalent $\tilde{\lambda}$, the values found for equivalent specific heats of channel and trench region are similar to the one of pure silicon. For instance, at 300 K c_{eq} evaluated for the channel model is only 5% bigger than the silicon value, while for the trench model c_{eq} is only 9% bigger than c_{Si}.

In order to verify the validity of the proposed assessments, transient thermal simulations have been performed. The verification consists of running a transient analysis using two different FE models. The first model, called *detailed* model, includes real material properties, while the second model, called *equivalent* model, features the same FE model of the detailed model, but it is made of only one single (homogeneous) material which is described by equivalent material properties computed with Equation 3.4, Equation 3.10 and Equation 3.11. An example of detailed and equivalent models is depicted in Figure 3.9 for the trench region. The transient simulation consist of applying a power pulse to a generation surface of the model and fixing the ambient temperature to the opposite sink surface along a certain direction η, similarly to the method described in Paragraph 3.2.1. For a given ambient temperature, the same simulation is performed for both the detailed and the equivalent FE model. Subsequently, both simulation results have been utilized for plotting and comparing thermal step responses (i.e. ΔT / time curves) of the two models. As examples, Figure 3.10 and Figure 3.11 show the thermal step responses for the detailed and the equivalent model of the trench region at 300 K along

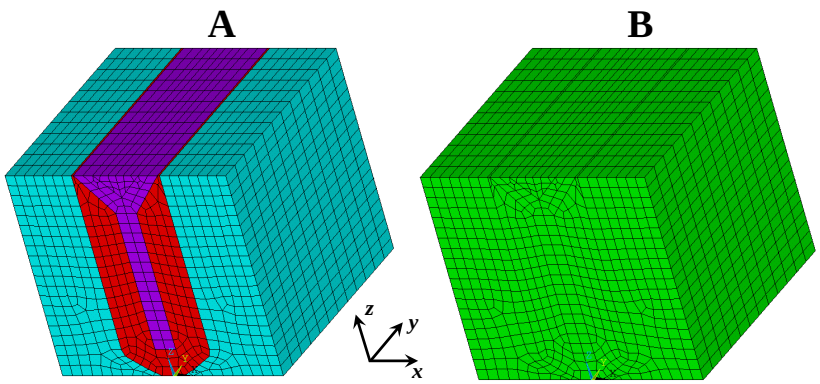

Figure 3.9: Detailed (A) and equivalent (B) FE model of the trench region.

the x and the z directions, respectively. The good agreement shown in the charts has been obtained also for different ambient temperatures as well as in the case of the channel region. These charts represent also valid verifications of orthotropic equivalent thermal conductivity computed by homogenization. In fact, the correspondence of values in the two thermal step responses indicates the equivalence of the effective λ (which is associated to the detailed model) to the one obtained by homogenization (which is associated to the equivalent model).

The thermal response comparison is particularly interesting along the x direction (see Figure 3.10), i.e. the direction along which the thermal barrier effect associated to the trench is relevant. This chart displays also the thermal step response of a homogeneous trench model (simple model), i.e. a FE model of the trench region composed of only silicon as in the case of a commonly adopted modeling of the epitaxial layer. The three displayed curves can be consistently compared because they have been obtained using the same power pulse. In Figure 3.10 the reader can notice a substantial difference between the thermal impedance of the pure silicon (used in the simple FE modeling of the epitaxial region) and the equivalent thermal impedance of the detailed trench region. Such discrepancy is visible over multiple time decades and it is due to the contribution of the substantially larger thermal impedances of trench oxides to the equivalent thermal impedance. Therefore, the outcome of modeling trench structures within the epitaxial layer is represented not only by a lowering of the thermal conductivity, but also by an increased thermal

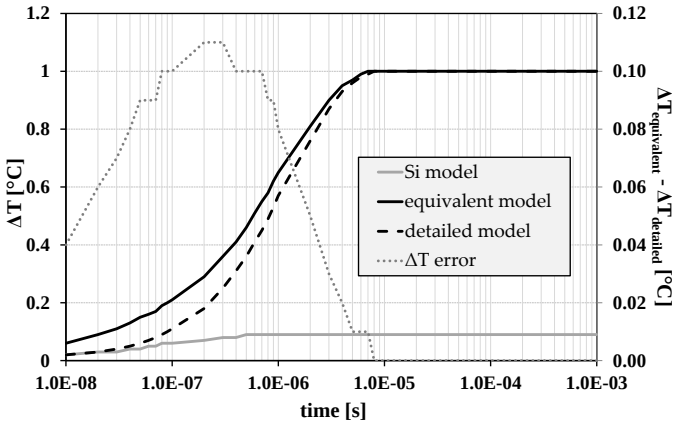

Figure 3.10: Simulated thermal step excitation for different trench models at 300 K, along x direction, by same superficial heat pulse. Dashed grey curve represents the temperature deviation between the equivalent and the detailed model.

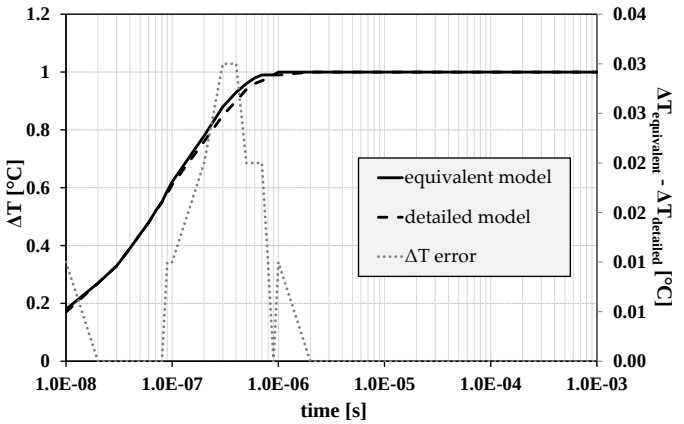

Figure 3.11: Simulated thermal step excitation for detailed and equivalent trench model at 300 K, along z direction, by same superficial heat pulse. Dashed grey curve represents the temperature deviation between the equivalent and the detailed model.

capacity. The first effect is obviously permanent because it is related to the steady state thermal behavior of the system (according to the definition of λ), while the latter effect vanishes after about 10 μs. Such an observation is very important because it will be useful for explaining the impact of the proposed device FE modeling on its electro-thermal behavior in Paragraph 3.2.6.

In this paragraph, an approach for assessing the equivalent thermal impedance associated to the detailed FE model of an elementary micro-structure has been presented and validated within the homogenization scenario. The advantage of using this approach is given by its relative simplicity, its effectiveness and its consistency with the physics background. Other methods have also been proposed in literature for evaluating equivalent thermal impedances of micro-models. In particular, an interesting method has been presented in [22, 83] where equivalent material parameters are evaluated within an optimization scheme with the goal of accurately matching the equivalent thermal response with the detailed model response. This method is actually quite effective because it is characterized by a good accuracy. Nevertheless, it cannot be considered completely consistent with the physical background and its implementation effort is definitely higher if compared with the simpler approach described in this thesis. Finally, equivalent material properties can also be assessed by exploiting a quite complex analytical formulation which was developed for solving inverse heat transfer problems [102].

3.2.5 Electrical modeling

With the homogenization procedure presented in the previous paragraphs, the equivalent thermal parameters associated to the trench micro-structure pattern within the epitaxial layer have been estimated. In order to perform electro-thermal simulations of a power MOSFET, electrical parameters of its representing FE model must be defined as well. In this paragraph the electrical modeling of the epitaxial layer stack presented in Paragraph 3.2.2 will be described. The reader should keep in mind that in the epitaxial region of a power device:

- Resistive contributions to the total device resistance are the largest [54];
- Power contributions to the total power dissipated within the device are the biggest;
- The switching mechanism of the semiconductor device takes place.

Therefore, the goal of electrical modeling is to accurately represent the electrical behavior of the epitaxial layer within the power device, thus such modeling has been developed based on physical considerations. Electrical material properties of the layers included in the epitaxial region have been defined as follow:

- In the channel layer, according to the fact that the current flows only vertically, the electrical resistivity associated to its generic finite element j has been set to an infinite value along the *x* and *y* directions, while the resistivity along the vertical direction, i.e. ρ_z, has been defined as follow:

$$\begin{cases} \rho_{x_j} = \infty \\ \rho_{y_j} = \infty \\ \rho_{z_j} = \frac{A_{\mathrm{j}}}{d_{\mathrm{j}}} \frac{\Delta V_{\mathrm{ch}_j}}{I_{\mathrm{ch}_j}(T_j, V_{\mathrm{G}_j})} \end{cases} \tag{3.13}$$

  The equations above define the orthotropic electrical behavior of the device channel within the implemented electro-thermal simulation algorithm. The expression for ρ_z has been already described in Paragraph 2.3.1 - Equation 2.58, while the temperature dependence of I_{ch} is due to the combined effect of the temperature dependent mobility in the channel and the temperature dependent threshold voltage (seen in Paragraph 2.1.1, Equation 2.1, Equation 2.2 and Equation 2.3).

- In the trench layer, only vertical conduction is assumed, hence an orthotropic resistivity has been set as in the case of the channel region, namely ρ_x and ρ_y are infinite, while ρ_z is given by:

$$\begin{cases} \rho_{x_j} = \infty \\ \rho_{y_j} = \infty \\ \rho_{z_j} = \frac{1}{q\overline{N_{\mathrm{D}}}\overline{\mu_{\mathrm{n}}}(\overline{N_{\mathrm{D}}}, T)} \end{cases} \tag{3.14}$$

  where $\overline{N_{\mathrm{D}}}$ represents the average donor doping level in the trench layer, while μ_{n} is the temperature and doping dependent electron mobility for which the model proposed by Reggiani et al. has been used [93]. Therefore, according to the electro-thermal interaction, the resistivity of the trench is also temperature dependent by means of the average mobility defined in this region.

- In the drift layer, the resistivity is defined as isotropic, temperature and doping dependent as in the modeling of the trench region, i.e. :

$$\rho_{x_j} = \rho_{y_j} = \rho_{z_j} = \frac{1}{q\overline{N_{\mathrm{D}}}\overline{\mu_{\mathrm{n}}}(\overline{N_{\mathrm{D}}}, T)} \tag{3.15}$$

The electrical modeling proposed above does not consider two important contributions to the total resistance associated to the epitaxial region of a power MOSFET, namely the *JFET resistance* and the *accumulation resistance*. In this case, the first contribution was neglected according to the fact that in a trench technology, contrary to a VD-MOSFET, the JFET region does not exist. On the other hand, as in the case of a VD-MOSFET, in a trench technology the formation of an electron accumulation layer takes place due to the spreading of the electron current at the trench base because of the positive gate bias. Generally, in a trench MOSFET, the ohmic contribution associated to such an accumulation layer has an impact on the total epitaxial resistance of at least one order of magnitude lower than resistive contributions associated to the channel and the drift regions [54], therefore the accumulation resistance has been neglected.

3.2.6 Impact on the device electro-thermal behavior

Two different FE models of the DUT1 (presented in Paragraph 2.4.1) have been used to perform electro-thermal simulations for evaluating the impact of the proposed epitaxial layer modeling. The first FE model features a commonly employed description of the epitaxial region, namely the epitaxial layer is geometrically described by means of a single layer of finite elements, where every element is characterized by silicon thermally isotropic material properties and by an orthotropic electrical resistivity defined according to Equation 2.58. This model has been conveniently called *SIMPLE model*. The second FE description of DUT1 is called *COMPLEX model* and its epitaxial region has been defined with four distinct layers of elements: one layer has been used for describing the drift region, two layers for the trench region and, finally, just one layer for the channel region. Equivalent material properties for those four layers have been defined according to the methodology presented above and are reported in Table 3.1 compared to material properties used for the SIMPLE model. In Figure 3.12 a cross-section of the FE model highlighting the layer structure is depicted for both the SIMPLE and the COMPLEX model.

In the following electro-thermal simulations, only one thermal boundary condition has been applied in order to model the heat sink. Such a condition

		COMPLEX		SIMPLE	
		λ	c	λ	c
Channel	x	29.0			
	y	59.1	744	74.3	709
	z	50.8			
Trench	x	6.4			
	y	52.1	783	74.3	709
	z	48.3			
Drift	x				
	y	74.3	709	74.3	709
	z				

Table 3.1: Thermally equivalent material properties for COMPLEX model sub-regions compared to isotropic silicon material properties (used for SIMPLE model) at 300 K. Measurement units according to SI: λ [W K^{-1} m^{-1}] and c [J kg^{-1} K^{-1}].

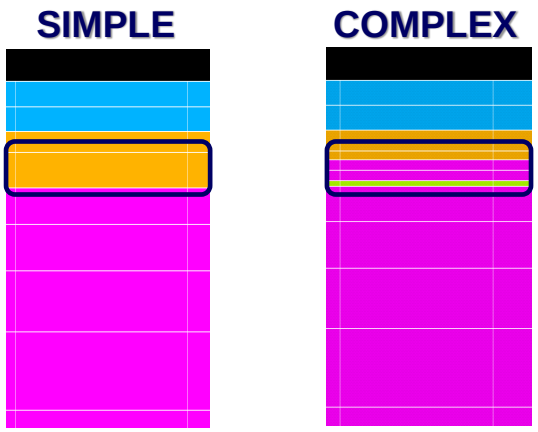

Figure 3.12: Cross-section of DUT1 FE model highlighting the layer structure for the SIMPLE and the COMPLEX model.

has been defined by fixing a low convective heat transfer at the chip bottom surface. All other external surfaces of the device models are set to be adiabatic.

Two different stress pulses have been simulated to evaluate the usefulness and the necessity of the proposed epitaxial layer modeling. The first stress condition is defined by a pulse with constant V_{GS} and with a duration of 1 ms. For

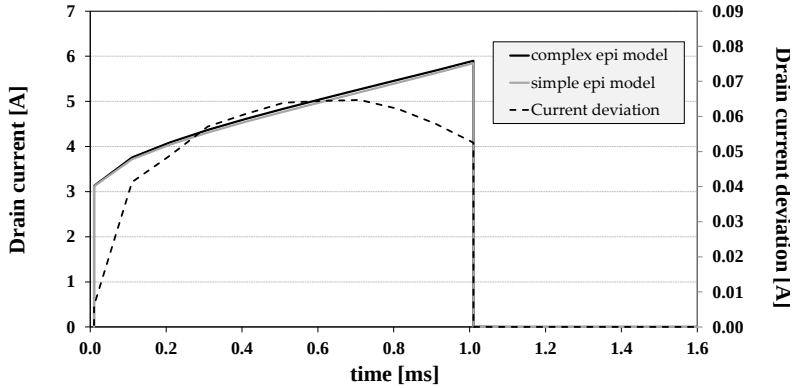

Figure 3.13: Drain current for the SIMPLE and the COMPLEX model for a low power/long pulse.

this pulse V_{GS} and V_{DS} are 2.2 V and 20 V, respectively. The corresponding power dissipation of this bias condition (i.e. about 90 W) is quite low when compared with a typical application overload situation. However, such a pulse defines a dangerous device operation because it implies a positive α which might cause the MOSFET to fail due to thermal runaway. Electro-thermal simulations of both the simple and the complex model have been performed using the algorithm described in Paragraph 2.3.1. Results show that at the end of the pulse, the relative deviation of the simulated drain current between the two epitaxial layer models is about 1% (see Figure 3.13). According to the very low current mismatch the simulated hottest point temperatures (depicted in Figure 3.14) also show a very low deviation, i.e. 3.5% at the end of the pulse. Such a small deviation is probably due to the relatively small volumes occupied by the oxide and BPSG materials within the channel and the trench geometries (as it can be seen in Figure 3.5), and because of the long timescale of the simulated stress pulse. The temperature distributions at the end of the pulse within the epitaxial layer for both epitaxial models (Figure 3.15) also show a good spatial matching. Furthermore, the reader should notice that the temperature distribution is not uniform because of the unstable thermal operation imposed by the simulated biasing point.

The second stress condition implies a substantially higher power dissipation in the device. Here DUT1 has been biased with a V_{GS} pulse of 5 V with

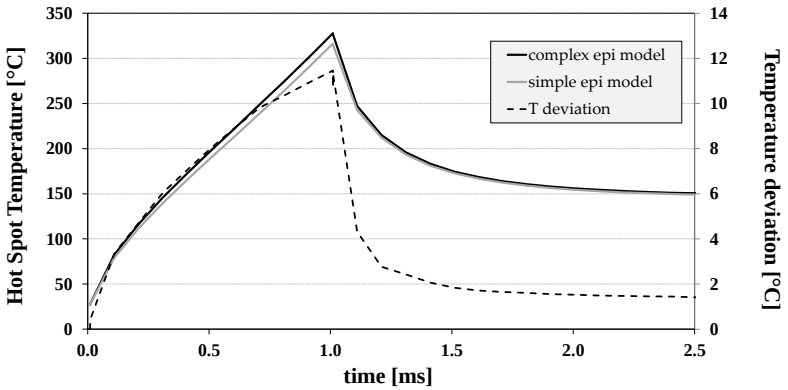

Figure 3.14: Hot-spot temperature for the SIMPLE and the COMPLEX model for a low power/long pulse.

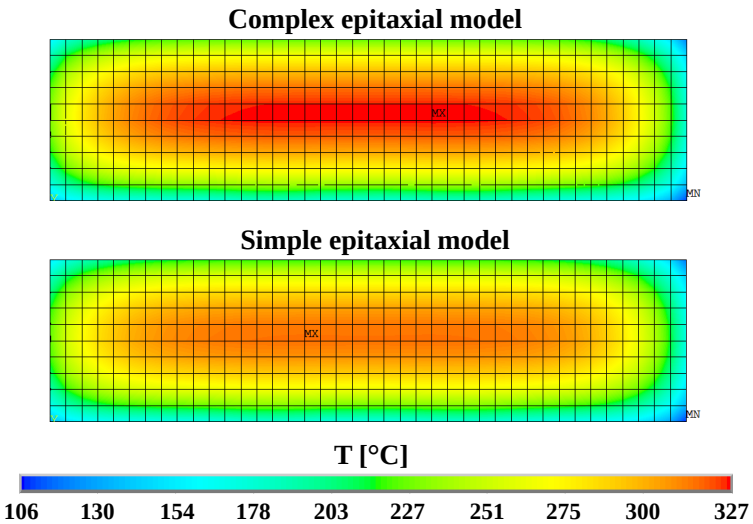

Figure 3.15: Temperature distribution within the epitaxial layer at the end of the low power/long pulse for the COMPLEX and the SIMPLE model.

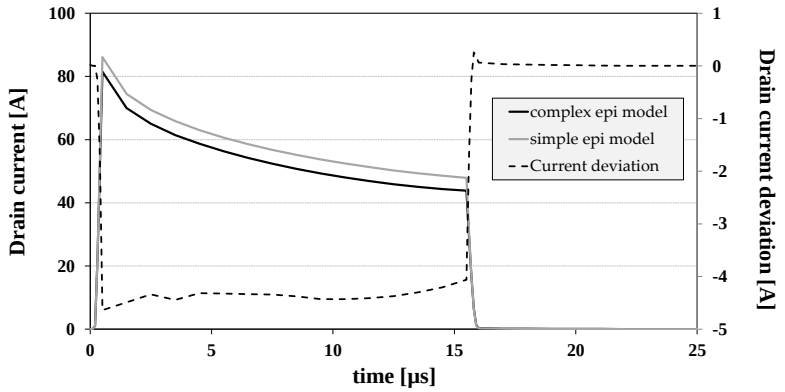

Figure 3.16: Drain current for the SIMPLE and the COMPLEX model for a high power/short pulse.

a constant V_{DS} of 40 V for a very short time, i.e. 15 µs. This biasing condition implies a negative α, therefore a thermally stable operation for the device. With this operating point the noticed deviations in the drain currents (reported in Figure 3.16) and in the hot spot temperatures (Figure 3.17) are more relevant, indicating the effect of the presence of the trench structures within the epitaxial layer. These deviations are mainly due to the contribution of thermal material properties during fast transients. Thermal simulations performed with the equivalent and the detailed models of both the trench and the channel layers show that after about 10 µs the dynamic thermal response is mainly dominated by $\tilde{\lambda}$ while the thermal capacity of the equivalent layer has no influence anymore, similar to what it is shown in Figure 3.10. There, the substantial difference between the thermal impedance of silicon (used in the simple model) and the equivalent thermal impedance of the complex model becomes visible in different time decades where the contribution of thermal capacities to thermal impedances finally vanishes. Therefore, power pulses acting in the time scale from 1 µs to 10 µs reveal a clearly visible thermal effect of the trench structures upon the device behavior. As it can be seen in Figure 3.14 and Figure 3.17, the complex model always returns hotter temperatures than the simple model because of the thermal considerations discussed above. Concerning drain current deviations between the two epitaxial models, in Figure 3.13 the device described with the complex model conducts slightly higher current than the

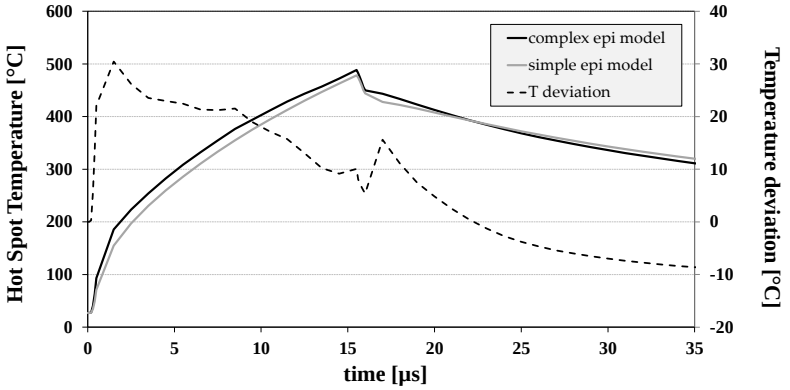

Figure 3.17: Hot-spot temperature for the SIMPLE and the COMPLEX model for a high power/short pulse.

device described with the simple one because the bias point implies a positive α. On the contrary, in Figure 3.16 the situation is reversed because the hotter device (i.e. the complex model) conducts less current than the simple model due to a negative α.

Finally, a few considerations concerning the computational time required to run the electro-thermal simulations: in the equivalent layer approach the FE model of the device consists of about $105 \cdot 10^3$ nodes, while in the simple approach only $85 \cdot 10^3$ nodes are needed. This difference is due to the fact that the device with the complex epitaxial model has three layers of finite elements (in the epitaxial stack) more than the device featuring the simple model. Those two layers imply an increase of the computational time needed to perform a single transient electro-thermal simulation by 40% - 50%. Therefore, together with the modeling effort required for the assessment of the equivalent material properties, the higher computational time needed by electro-thermal simulations using the equivalent model represents the main limitation of the proposed method.

Chapter 4

Applications

When assessing the robustness and/or the reliability of power devices, such as power MOSFETs or IGBTs, a thermal simulator generally represents a valid tool for qualitatively predicting the device behavior. During critical operations, like a SC event, the device junction may easily reach very high peak temperatures. Furthermore, during particular operations, the device may be exposed to large thermal gradients even within its small (a few square millimeters) active region, as in the case of a thermally unstable operating point. Temperature peaks and thermal gradients are main driving causes which lead to device degradation and failures, due to the fact that those thermal effects directly interact with the device mechanical and electrical behavior. They generally imply mechanical stresses due to thermal expansions associated to device compound materials, e.g. silicon, metal, oxide, etc.. These materials show a significantly different Coefficient of Thermal Expansion (CTE), therefore during a quick temperature rise, for instance associated to a SC, thermal expansions of metal, silicon, oxide and mold compound may be substantially different, causing mechanical stresses which progressively degrade the device integrity. On the other hand, extreme thermal excursions also determine substantial variations of device parameters which, in turn, determine the stability/instability behavior of the power switch. Therefore, when predicting the power device performance in critical operations, a purely thermal simulator should be used very carefully. More recommended is rather the usage of an electro-thermal or a thermo-mechanical simulator depending on the issue that has to be investigated.

This chapter is devoted to demonstrating that the proposed electro-thermal simulator is a useful tool that can be consistently employed for several kinds

of analyses aimed at the estimation of device reliability and, especially, robustness. This tool is very useful for designers particularly when analyzing causes of degradations (or failures) in an already existing technology or during the development of new ones.

Two applications of scientific relevance will be shown in this chapter. The first one concerns the impact on the device robustness of the progressively increasing current capability trend in latest Smart Power DMOS. The second one is a qualitative study about the electro-thermal impact of a growing crack within a copper-based power metallization, which represents a typical effect observed in power MOSFETs subjected to repetitive SC pulses.

4.1 Technology impact on device thermal stability

4.1.1 DMOS trends in modern Smart Power switches

In last decades, power MOSFETs integrated in SP switches have known two distinct shrinking trends, the first one is related to their active area size, while the latter concerns their R_{ON} parameter.

The active area downsizing seen in latest power MOSFET devices is mainly due to the increasing demand for smaller electronic components in modern (and future) automotive applications. In fact, a smaller active region allows the usage of smaller chip and package geometries. Furthermore, a smaller DMOS releases further space which could be used for the "SMART" analog circuitry integrated within the SP switch.

The ON-resistance R_{ON} of a power MOSFET is defined as the total electrical resistance to current flow between the drain and the source terminals at positive gate biasing potential. This parameter is of fundamental importance because it limits the power capability of the device. The power dissipated in the MOSFET in the ON-state operation is given by:

$$P_{\mathrm{diss}} = V_{\mathrm{DS}} I_{\mathrm{D}} = R_{\mathrm{ON}} I_{\mathrm{D}}^2 \tag{4.1}$$

In the expression above, for the same current level, a higher R_{ON} leads to a higher power dissipation within the device. Generally, R_{ON} can be described by the sum of several resistive contributions:

$$R_{\mathrm{ON}} = R_{\mathrm{S_{cont}}} + R_{\mathrm{S_{Si}}} + R_{\mathrm{ch}} + R_{\mathrm{acc}} + R_{\mathrm{JFET}} + R_{\mathrm{drift}} + R_{\mathrm{sub}} + R_{\mathrm{D_{Si}}} \tag{4.2}$$

where:

- $R_{\mathrm{S_{cont}}}$ is the source contact resistance between the n^+ source region and its corresponding electrode;
- $R_{\mathrm{S_{Si}}}$ is the resistance associated to the n^+ silicon defining the source region;
- R_{ch} is the device channel resistance, which is defined in the linear regime of operations as follows:

$$R_{\mathrm{ch}} = \frac{V_{\mathrm{DS}}}{I_{\mathrm{D}}} = \frac{L_{\mathrm{ch}}}{\mu_{\mathrm{n}} C_{\mathrm{ox}} W_{\mathrm{eff}} (V_{\mathrm{GS}} - V_{\mathrm{th}})} = \frac{1}{2K_{\mathrm{MOS}}(V_{\mathrm{GS}} - V_{\mathrm{th}})} \quad (4.3)$$

  where K_{MOS} is the transconductance coefficient (or conductivity parameter) of the MOSFET given by:

$$K_{\mathrm{MOS}} = \frac{1}{2}\mu_{\mathrm{n}} C_{\mathrm{OX}} \frac{W_{\mathrm{eff}}}{L_{\mathrm{ch}}} = \frac{1}{2}\mu_{\mathrm{n}} \frac{\epsilon_{\mathrm{OX}}}{t_{\mathrm{OX}}} \frac{W_{\mathrm{eff}}}{L_{\mathrm{ch}}} \quad (4.4)$$

- R_{acc} is the accumulation resistance which models the formation of an electron accumulation layer during the current spreading at the end of the channel (beginning of the body region) because of the positive gate bias applied;
- R_{JFET} is the JFET resistance associated to the JFET region which can be observed between the channel and the drift region in a VD-MOSFET only;
- R_{drift} is the drift resistance resulting from the drift of the current flow through the low doped *n* silicon in the epitaxial region. This resistance contribution is determined by geometrical parameters, such as cell pitch dimension, trench length and width, mesa region width, junction depth of the *p*-base region, etc.;
- R_{sub} is the resistance associated to the silicon substrate;
- $R_{\mathrm{D_{Si}}}$ is the resistance associated to the n^+ silicon defining the drain region;

For a detailed description of those resistive contributions, the reader may refer to [54]. In a power Trench MOSFET, biggest contributions to R_{ON} are given by the channel resistance R_{ch} and the drift resistance R_{drift}, therefore substituting Equation 4.2 into Equation 4.1 we obtain:

$$P_{\mathrm{diss}} = R_{\mathrm{ON}} I_{\mathrm{D}}^2 \approx (R_{\mathrm{ch}} + R_{\mathrm{drift}}) I_{\mathrm{D}}^2 \quad (4.5)$$

A smaller active area implies a smaller MOSFET effective channel width W_{eff}, which is given by:

$$W_{\text{eff}} = \sum\nolimits_j W_j = NW_j \tag{4.6}$$

where W_j is the channel width of a single cell and N is the total number of cells. Consequently, the lower W_{eff}, the higher R_{ch}. On the other hand, smaller active areas determine also bigger R_{drift}. Therefore, the effect of such a shrink leads to a remarkable drawback, i.e. the device would dissipate higher power for the same current level due to an increased R_{ch} and R_{drift}. In other words, active area shrinking causes a reduced current capability in standard applications. In order to counteract this unfavourable drawback, the device active area shrinking is often accompanied by an increased device transconductance g_{m}, which is given by:

$$g_{\text{m}} = \frac{dI_{\text{D}}}{dV_{\text{G}}} = \begin{cases} \text{if } \; V_{\text{GS}} - V_{\text{th}} \geq V_{\text{DS}} : \\ \mu_{\text{n}} C_{\text{OX}} \frac{W_{\text{eff}}}{L_{\text{ch}}} V_{\text{DS}} = 2K_{\text{MOS}} V_{\text{DS}} \\ \\ \text{if } \; V_{\text{GS}} - V_{\text{th}} < V_{\text{DS}} : \\ \mu_{\text{n}} C_{\text{OX}} \frac{W_{\text{eff}}}{L_{\text{ch}}} \left(V_{\text{GS}} - V_{\text{th}}\right) = 2K_{\text{MOS}} \left(V_{\text{GS}} - V_{\text{th}}\right) \end{cases} \tag{4.7}$$

Nowadays, the way to reach higher device transconductance in power MOSFETs is generally achieved by the introduction of consistent technology improvements which often result in a higher K_{MOS}, for instance by shortening the channel length L_{ch} or by narrowing the gate oxide thickness t_{OX}. More commonly, an effectively increased K_{MOS} can be obtained by increasing the trench density introducing an opportune technology scaling [103, 104, 105, 106], which implies a smaller pitch size and consequently a larger W_{eff} because of a greater number of cells. For the same active area a smaller pitch determines indeed an effectively increased device current capability due to a larger number of source to drain conduction paths. In general, the higher the K_{MOS}, the lower the R_{ON} because of smaller R_{ch}.

4.1.2 Impact of the active area size on the robustness

In the previous paragraph it was shown that the problem of the reduced current capability related to the miniaturization trend is actually counteracted by achieving a technology featuring a higher transconductance. Nevertheless, the downsizing of the DMOS area still creates a relevant issue related to device

robustness. In fact, a smaller active area determines a higher temperature peak within the device due to an increased junction to case thermal impedance $Z_{\mathrm{th_{j-c}}}$. The $Z_{\mathrm{th_{j-c}}}$ parameter defines the temperature rise $\Delta T_{\mathrm{j-c}}$, i.e. the difference between the junction temperature and the case temperature, for the power dissipation P_{diss}:

$$\Delta T_{\mathrm{j-c}} = T_{\mathrm{j}} - T_{\mathrm{c}} = Z_{\mathrm{th_{j-c}}} P_{\mathrm{diss}} \tag{4.8}$$

For example, in a very simplified case (short power pulse on top of a semi-infinite plate) the temperature peak can be estimated by means of the "square root law" proposed by Glavanovics and Zitta [107]:

$$\Delta T_{\mathrm{j-c}} = \frac{P_{\mathrm{diss}}}{A_{\mathrm{DMOS}}} k_{\mathrm{Si}} \sqrt{t_{\mathrm{pulse}}} \tag{4.9}$$

where k_{Si} represents the "root law factor" which depends on silicon thermal material properties. For a given power pulse of a given duration, Equation 4.9 well shows the trade-off between junction temperature peak and DMOS active area. Intuitively, from a physical point of view, it is clear that a smaller active area determines a larger $Z_{\mathrm{th_{j-c}}}$ due to: 1) a smaller are available for thermal exchange between the epitaxial layer and heat sink; 2) a reduced thermal capacity because of a smaller silicon mass involved in the heat exchange. On the other hand, Alatise et al. [24] have shown experimentally that, aside from the aforementioned purely thermal considerations, the active area shrinking would generally shift the V_{TCP} towards lower V_{GS} values, with a consequently improved device thermal stability and robustness behavior.

In conclusion, the active area shrinking leads to two counteracting effects on the device robustness, namely an unfavorable one given by a larger thermal impedance and a favorable one represented by a larger thermal stability range. In order to face the problem of the reduced current capability, higher K_{MOS} has been achieved in latest power MOSFET technologies, therefore possible robustness issues related to such evolution should be addressed as well.

4.1.3 Impact of the K_{MOS} coefficient on the robustness

The impact of the K_{MOS} coefficient on the device electro-thermal behavior has already been analytically and experimentally described in literature [12, 24, 13], though never verified by simulations. In this paragraph, the trade-off between K_{MOS} and robustness will be demonstrated by performing electro-thermal simulations of two test chips. Both test devices are based on the DUT2 already presented in Paragraph 2.4.1.

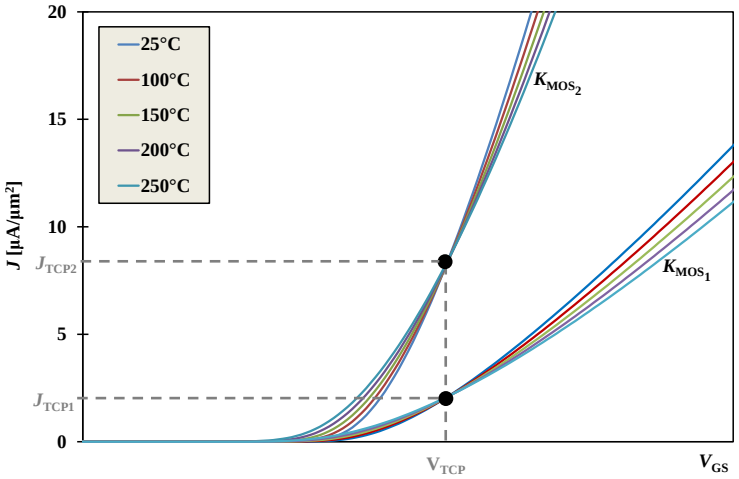

Figure 4.1: Transfer characteristics at different ambient temperatures for the two technologies. As expected, the current density level at the TCP is larger in technology-2.

Two versions of the DUT2 have been used for this investigation, both test devices have the same chip size, active area dimensions, number of bond wires and geometries and they both incorporate within their respective active region (at the same relative position) an integrated bipolar temperature sensor. The temperature sensor has been employed for validating simulation results achieved by using the algorithm for simulating SP switches described in Paragraph 2.3.3. The reader can find details about the temperature sensor and its calibration in Paragraph 2.4.3, while the experimental validation has been already treated in Paragraph 2.4.4. The first test chip (DUT1-A) has been fabricated using a trench technology (technology-1) characterized by the transconductance coefficient value K_{MOS_1}, while the second one (DUTB-2) features a similar trench technology (technology-2) showing a larger current capability and, therefore, a bigger value of the transconductance coefficient (i.e. K_{MOS_2}).

From the plot of the two transfer characteristics associated to the two technologies (Figure 4.1) it can be seen that the two DUTs share the same voltage value at the TCP, i.e. their V_{TCP} is the same, but the current value which defines the onset of the thermal instability in technology-2 is higher ($I_{\mathrm{TCP2}} > I_{\mathrm{TCP1}}$). This means that technology-2 is characterized by a larger instability range compared to technology-1. Such observation can be verified by deducing the

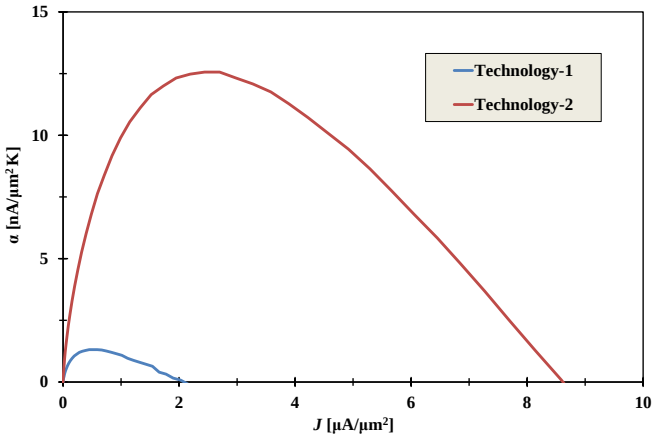

Figure 4.2: Drain current temperature coefficient α as a function of the current density for both technologies. As expected, both the temperature coefficient peak and the current instability range are larger for the technology-2.

drain current temperature coefficient α as a function of the drain current value [12], i.e.:

$$\alpha(I_{\mathrm{D}}) = -\frac{M}{T} I_{\mathrm{D}} + 2\Phi\sqrt{K_{\mathrm{MOS}} I_{\mathrm{D}}} \tag{4.10}$$

In the equation above, T is the absolute temperature in Kelvin, Φ is the temperature coefficient of the threshold voltage V_{th} (see Equation 1.13) and M the exponential term of the temperature dependent mobility at low-fields in Equation 1.12.

In the chart depicted in Figure 4.2, Equation 4.10 has been plotted for both technologies scaled to current density. Here, it can be clearly seen that for the DUT2-B α is positive for a wider I_{D} range, furthermore also its α peak value (i.e. α_{max}) is bigger. Indeed, according to expressions derived by Spirito et al. [12]:

$$\alpha_{\mathrm{max}} = K_{\mathrm{MOS}} \frac{\Phi^2 T}{M} \tag{4.11}$$

$$I_{\mathrm{TCP}} = K_{\mathrm{MOS}} \left(\frac{2\Phi T}{M} \right)^2 \tag{4.12}$$

the higher K_{MOS}, the larger α_{max} and I_{TCP}.

Four different electrical stress conditions have been simulated using just one FE model. From the simulation point of view, the only difference between

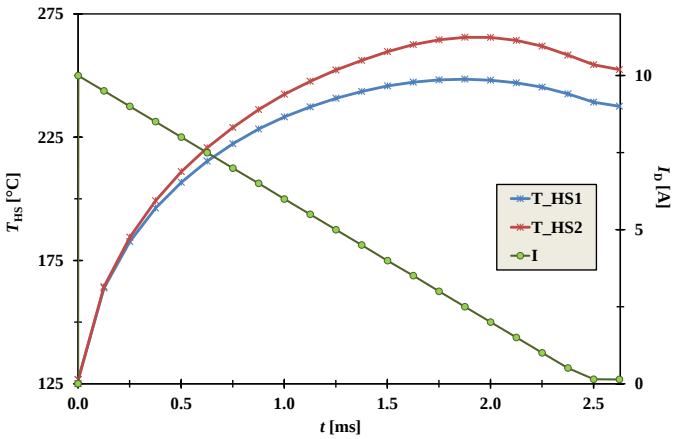

Figure 4.3: Simulated hot-spot temperature T_{HS} for I_{peak} = 10 A, V_{DS} = 40 V, T_{pulse} = 2.5 ms and T_{ambient} = 125°C. As expected $T_{\mathrm{HS2}} > T_{\mathrm{HS1}}$ because $\alpha_2 > \alpha_1 > 0$. Maximum temperature deviation between T_{HS2} and T_{HS1} is 17.3°C (at t = 2 ms).

DUT2-A and DUT2-B lies in the table model applied for the definition of the channel element electrical resistivity ρ_{ch_j} (Equation 2.58). Indeed, different g_{m} (or analogously different K_{MOS}) imply different drain current values for the same V_{GS} biasing level which finally defines different values for ρ_{ch_j}.

The first stress condition is described by a triangular power pulse which emulates the discharge of the magnetic energy stored into an inductive load. Analogous pulses have been already employed in Paragraph 2.4.4 for validating the simulator, where the measured/simulated sensor temperatures in the two test devices (see Figure 2.41 and Figure 2.42) do not exhibit significant differences because of the short transient employed (500 µs). The difference in the electro-thermal behavior rather arises for instance if a longer pulse is used, as in the case plotted in Figure 4.3 where the simulated hot-spot temperature T_{HS} is slightly higher (about 17°C) in DUT2-B. Such a temperature deviation is a direct consequence of the thermally unstable working regime defined for both devices by the stress pulse, in fact T_{HS} in DUT2-B is higher due to its larger α.

The second simulated stress pulse also implies a temperature unstable regime for both DUTs by defining a more severe operation. Here, the simulated transient consists of a constant power pulse of 400 W obtained using a square current pulse of 10 A at $V_{\mathrm{DS}} = 40V$. In Figure 4.4 the hot-spot tem-

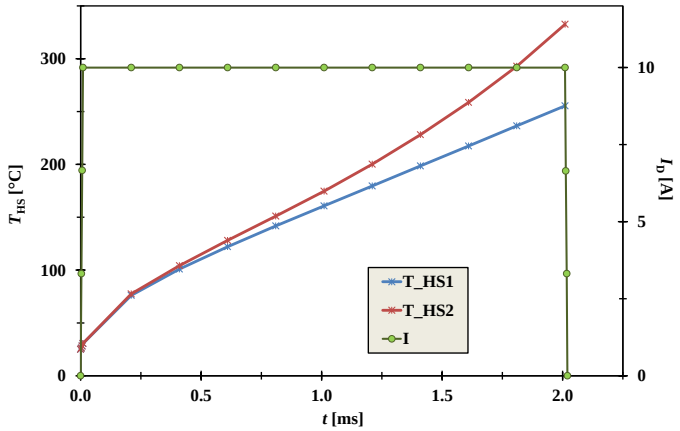

Figure 4.4: Simulated hot-spot temperature T_{HS} for I_{D} = 10 A, V_{DS} = 40 V, T_{pulse} = 2.0 ms and T_{ambient} = 25°C. A longer pulse would lead DUT2-B to fail earlier than DUT2-A because $T_{\mathrm{HS2}} > T_{\mathrm{HS1}}$. At the end of the pulse the temperature deviation between T_{HS2} and T_{HS1} is 77.1°C.

perature rising in DUTB-2 is substantially faster than in DUTB-1, in fact at the end of the 2 ms pulse the temperature deviation between T_{HS1} and T_{HS2} is already 77 K. Furthermore, due to electro-thermal interactions and to applied thermal boundary conditions, the temperature distribution in DUTB-2 is less homogeneous if compared with the one simulated in DUTB-1, as depicted in Figure 4.5. A longer pulse would lead, after a sufficient time, both devices to thermal runaway due to the focusing of a hot-spot associated with a localized current crowding [25]. Nevertheless, DUTB-2 will surely fail earlier (namely in a shorter time) due to its faster temperature rise.

The third simulated stress pulse is quite interesting because it implies different regimes for the two test devices: a thermally stable for DUTB-1 and a thermally unstable one for DUTB-2. In fact, the selected drain current value, i.e. I_{D} = 25 A, has been chosen bigger than I_{TCP1} and smaller than I_{TCP2}, therefore during the simulated biasing condition α_1 is always negative and α_2 is always positive. Simulation results reported in Figure 4.6 show substantially different thermal behavior of the two devices, in particular the very steep increase of T_{HS2} observable after about 1 ms indicates the occurrence of the thermal runaway phenomenon. Such a failure does not take place in the DUTB-1 even at the end of the 1.5 ms pulse, specifically after about 1 ms T_{HS1} is still well below 400°C.

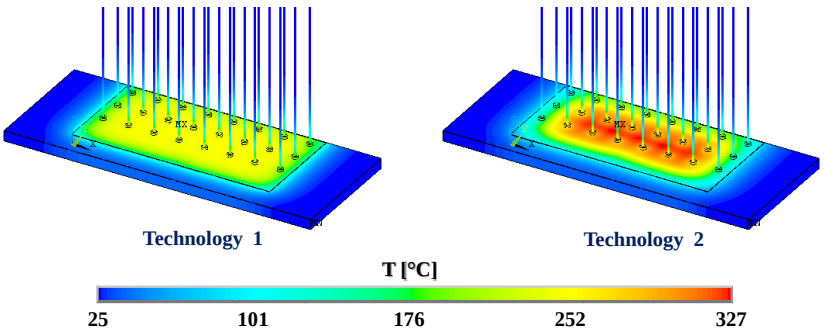

Figure 4.5: Simulated temperature distributions across DUT2-A and DUT2-B at the end of the current pulse of Figure 4.4. The test device realized with technology-2 shows a higher and more inhomogeneous temperature distribution than the one simulated using technology-1.

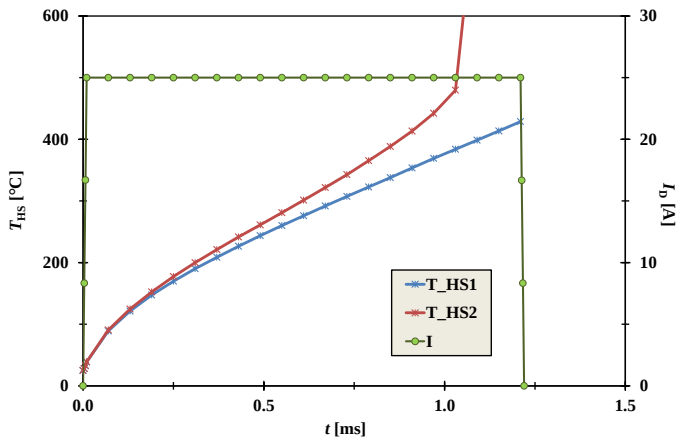

Figure 4.6: Simulated hot-spot temperature T_{HS} for $I_{\mathrm{D}} = 25$ A, $V_{\mathrm{DS}} = 40$ V, $T_{\mathrm{pulse}} = 1.2$ ms and $T_{\mathrm{ambient}} = 25°\mathrm{C}$. The biasing condition simulated here defines a positive α_2 and, at the same time, a negative α_1. The thermal runaway phenomenon can be observed in DUT2-B after about 1 ms due to its thermally unstable operation.

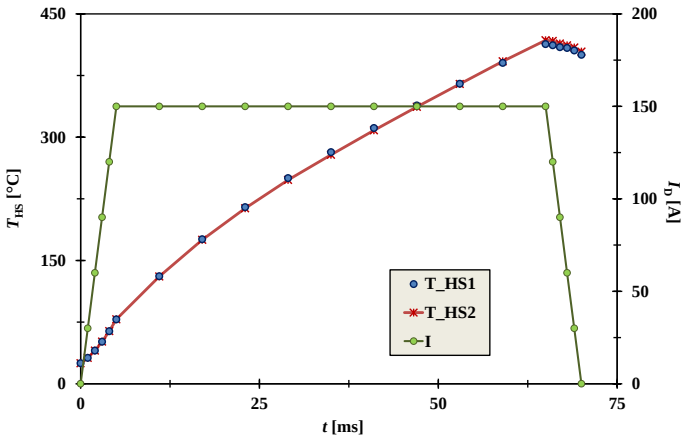

Figure 4.7: Simulated hot-spot temperature T_{HS} for I_{D} = 150 A, V_{DS} = 40 V, T_{pulse} = 60 µs and T_{ambient} = 25°C. Because of the thermally stable operations, the rise of T_{HS} in the time is the same for both DUTs indicating that about the TCP the difference between α_1 and α_2 does not have an influence on the device thermal response.

Finally, the last simulated pulse is defined by applying a large I_{D} value so that both devices work within their thermally stable regions. In Figure 4.7, both simulated temperature hot-spots experience exactly the same temporal evolution indicating that in this case the technology difference does not influence the electro-thermal behavior of the MOSFETs. This means that for thermally stable regimes, $\alpha_1 < \alpha_1 < 0$ does not imply a "more stable" operation for the DUT2-A, in other words the negative value of α does not change the device's electro-thermally stable behavior due to the negative feedback between the temperature and the drain current.

In conclusion, the faster rise of T_{HS2} compared to T_{HS1} observed in all the simulated thermally unstable conditions well demonstrates that the higher the K_{MOS} coefficient, the shorter the time required to drive the device to the thermal runaway. Therefore, such an issue should always be addressed when evaluating the robustness of new power MOSFET technologies, keeping in mind that (as already described in the Paragraph 1.2.3) electro-thermal instability further limits the SOA of the device at high-V_{DS}/low-I_{D} operations. Part of the results presented in this section will be published in the proceeding of the IEEE 25^{th} International Symposium on Power Semiconductor Devices and ICs (ISPSD 2013) [36].

4.2 Cracks within the source metallization

4.2.1 A reliability issue in actively cycled devices

The electro-thermal simulator presented in this thesis is particularly suitable for predicting device electrical and thermal behavior under a single stress condition (single pulse). Just one electro-thermal simulation may typically require several hours for computing fields solutions, depending on the FE model complexity (i.e. number of nodes/elements), on the number of time steps for describing the input waveforms, and on the employed solution algorithm (discrete device versus SP switch). Obviously, simulating a repetitive stress condition (cycling pulse) is also possible, but it would normally imply an unacceptable computational time for calculating the electrical and thermal field distributions. Based on the robustness and reliability concepts described in Paragraph 1.1.3, it may be stated that this electro-thermal simulator is particularly suitable for evaluating the device's robustness characteristic, rather than its reliability. Nevertheless, in this paragraph it will be shown how such a tool can be also used for assessing a typical reliability issue of modern SP switches.

During their life time, SP devices have to reliably withstand up to millions of stress power pulses, such as SC. In this case, fatigue mechanisms have to be considered, in fact due to the repeated thermo-mechanical strain induced, the power metallization and the interconnection stack may rapidly degrade and lose their initial integrity properties [108, 6, 109, 110]. Such degradation mechanisms may drive the generation of a local hot-spot within the device, which may lead the device to failure due to thermal runaway after a sufficient number of stress pulses.

Nowadays, the DMOS integrated in modern SP switches is capable of handling millions of SC pulses without failure even if the peak temperature induced in the silicon by a single pulse can easily reach temperatures higher than 300°C. This is possible thanks to the temperature margin ΔT_j existing between initial temperature peak T_j^{peak} of the virgin device and the destruction temperature limit T_j^{SOA}, where the latter limit is directly related to the device robustness. During repetitive SC test, every power pulse induces a thermo-mechanical stress, i.e. plastic deformation, on the device metallization and interconnections which results in a higher and higher temperature peak T_j^{peak} within the silicon. As depicted in Figure 4.8, after every cycle, due to this degradation mechanism, the headroom between T_j^{peak} and T_j^{SOA} progressively diminishes. Improving the device reliability means also realizing a DMOS in which ΔT_j degrades slowly, achieving a higher number of cycles to failure.

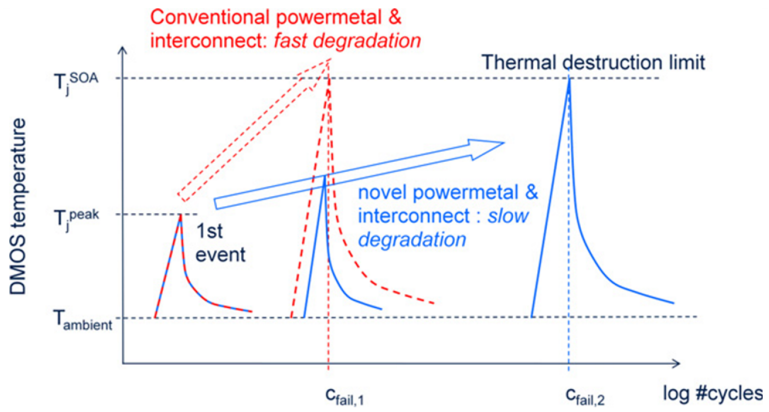

Figure 4.8: Scheme describing device degradation in active cycle stress: the novel copper-based technology, as opposed to a conventional aluminum-based technology, leads to a slowly growing peak temperature in the silicon $T_{\mathrm{j}}^{\mathrm{peak}}$ due to a less severe degradation of the power metal and interconnections, allowing a higher destruction limit, i.e. $c_{\mathrm{fail},2} > c_{\mathrm{fail},1}$ [10].

Nelhiebel at al. [10] have developed and presented a very reliable technology concept characterized by a copper-based power metallization, instead of the commonly used aluminum, which allows substantially higher cycles to failure, hence a higher MOSFET reliability.

Experimental results and failure analyses have demonstrated that in this technology the failure mechanism is related to the formation of cracks and voids within the copper metallization in the proximity of the ILD (see FIB micrograph in Figure 4.9) due to the plastic deformation fatigue induced by every pulse. Tests conducted on a big population of specimens have shown that after a sufficient number of pulses (about 10 millions) cracks and voids start to emerge and grow within the metallization, while first runaway-fails are observed around 20 million pulses due to some severe localized cracks/voids in the copper.

4.2.2 Electro-thermal effects of cracks

In this paragraph the impact of a crack in the power metallization will be analyzed by means of electro-thermal simulations.

The electro-thermal effect of such a crack is not obvious because, from the thermal point of view, it entails a less effective heat extraction from the sili-

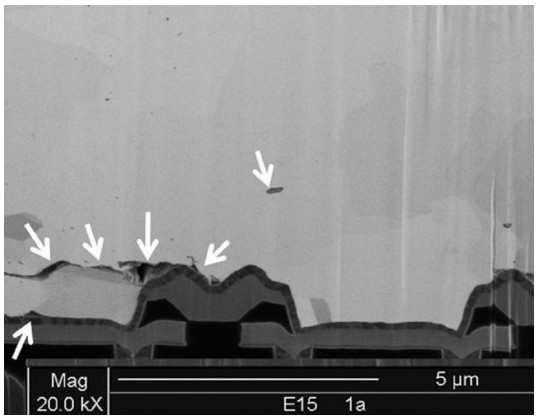

Figure 4.9: FIB micrograph of voids and cracks observed in the copper power metallization [10].

con due to the reduced heat capacity and thermal conductivity associated to the void, while from the electrical point of view, it implies a smaller (local) current due to a "de-biased" source potential. In fact, it is reasonable to suppose that the source potential just below a crack is slightly higher than in the undisturbed case. Therefore, the local V_S seen by a cell placed just under a crack is higher, hence its local current is lower because of a diminished V_{GS}. Finally, a crack in the power metallization defines two opposite thermal effects: a local increase of temperature related to an increased thermal impedance associated to the cracked-metal and a local decrease of temperature associated to the lower dissipated power thanks to the reduced cell current and effective V_{DS}.

The FE model of DUT1 introduced in Paragraph 2.4.1 and depicted in Figure 2.13 has been slightly modified towards a finer mesh. The adopted mesh has the following characteristics:

- Every square element of the epitaxial layer includes 4 (2 x 2) elementary cells in the *xy* plane, i.e. the element size along *x* and *y* directions is 12.8 µm;
- For the power metallization two element layers have been used, i.e. two elements have been employed along the *z* direction.

With such a meshing strategy, a crack has been modeled by defining four or more *void elements* in the lower power metal layer. Four different void lengths have been simulated:

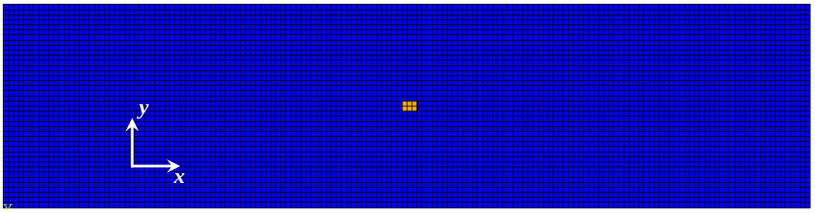

Figure 4.10: Example of a 38.4 µm crack (3 x 2 orange elements) placed on the top of the the innermost area of the active region. In the adopted mesh, every element of the device active area (blue elements) features 2 x 2 elementary cells (12.8 µm x 12.8 µm).

- 25.6 µm, by using 2 x 2 void-elements, i.e. the whole void is a square: $2 \cdot 12.8$ µm along *x* and $2 \cdot 12.8$ µm along *y*;
- 38.4 µm, by using 3 x 2 void-elements, i.e. the whole void is a rectangle: $3 \cdot 12.8$ µm along *x* and $2 \cdot 12.8$ µm along *y*;
- 51.2 µm, by using 4 x 2 void-elements, i.e. the whole void is a rectangle: $4 \cdot 12.8$ µm along *x* and $2 \cdot 12.8$ µm along *y*;
- 64.0 µm, by using 5 x 2 void-elements, i.e. the whole void is a rectangle: $5 \cdot 12.8$ µm along *x* and $2 \cdot 12.8$ µm along *y*.

From the thermal and electrical point of view a crack behaves as a thermal and electrical isolator, hence a void-element features a very high electrical resistivity, very small thermal conductivity and thermal capacity values.

Just one single void has been placed at the hottest site of the active region, i.e. the central innermost area in the *xy* plane. For example, the position of a 3 x 2 crack, together with the mesh adopted for the epitaxial layer is visible in Figure 4.10.

The electro-thermal effect of a crack has been estimated by simulating a typical SP application pulse emulating the fast discharge of the magnetic energy stored within an inductive load, namely a triangular current shape with $I_{\mathrm{peak}} = 65\mathrm{A}$, constant V_{DS} to 41.5 V and pulse duration of 25 µs. The same pulse has been used for simulating the device behavior for all crack sizes described above. Simulation results show that the device temperature hot-spot T_{HS} emerges in a location of the epitaxial region situated just below the metal void. Furthermore, the plot in Figure 4.11 shows that T_{HS} increases for increasing crack sizes, indicating that the higher thermal impedance associated

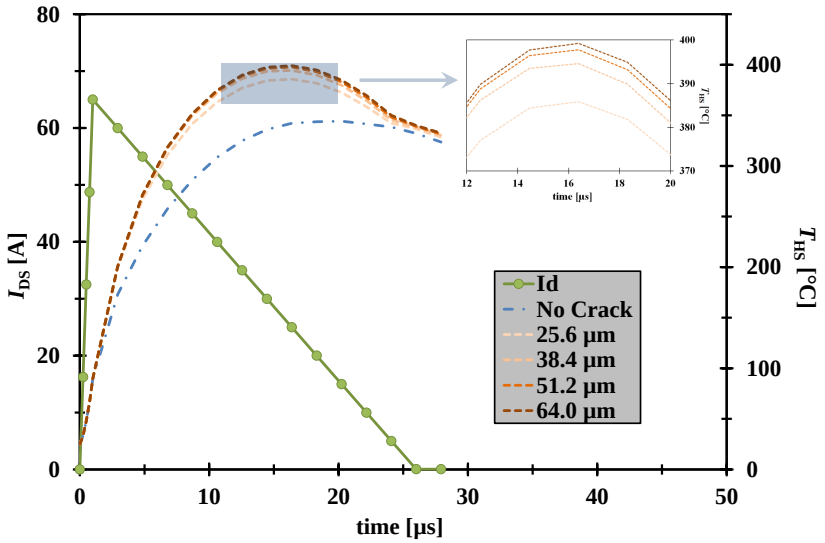

Figure 4.11: Simulated drain current (green curve) and hot-spot temperature for the free-crack model (blue dash-point curve) and for different crack lengths (orange-shades dashed curve).

to the crack acts like a thermal barrier for the evacuation of the heat generated within the epitaxial silicon.

For a better understanding, in Figure 4.12 the hot-spot temperature T_{HS} and the temperature overhead (with respect to the free-crack model) ΔT_{HS} have been plotted as a function of the crack length. For cracks longer than 25 µm the hot-spot temperature overhead ΔT_{HS} is larger than 40 K. For larger crack sizes both T_{HS} and ΔT_{HS} grow slower and slower, resembling a sort of saturation effect, which could be explained considering the electro-thermal coupling that takes place within a device cell. In the insert of Figure 4.11, it can be seen that the hot-spot temperature peak is reached in the cracked-models after about 16 µs, i.e. when the drain current is about 25 A. This current level is actually well about the I_{TCP} level, indicating a temperature stable regime for the device, hence a negative α. This means that, starting from a certain crack size, the temperature increase in the silicon caused by the thermal barrier is counterbalanced by a local current decrease which, in turn, defines a smaller power. Such supposition is partly confirmed in Figure 4.13, where both the temperature and the current density fields in the active region have been depicted for the 51.2 µm crack. There, the reader should notice that:

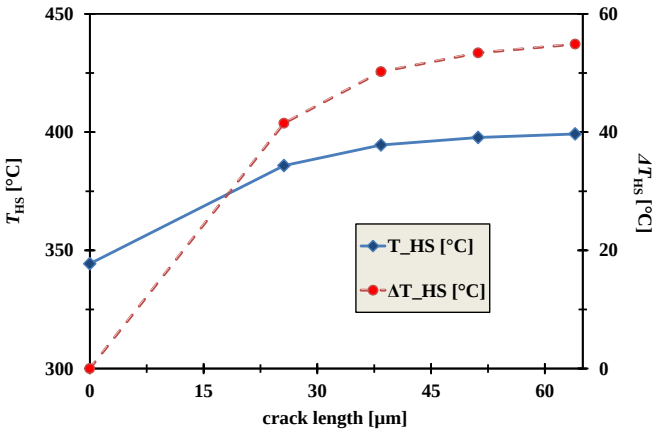

Figure 4.12: Temperature hot-spot T_{HS} (blue curve) and temperature hot-spot overhead (respect to a free-crack device) ΔT_{HS} (red curve) for different crack lengths.

1) the temperature hot-spot in the epitaxial layer emerges exactly under the power metal void; 2) the current density in the active area is lower just at the location where the temperature is higher, i.e. again in the silicon under the crack.

In conclusion, the simplified analysis presented in this paragraph describes only one possible reason for the device failure shown in [10]. In particular, contrary to what has been experimentally observed, here just one single crack within the device metallization has been modeled and simulated by FEM. Simulation results have shown that a large crack in the power metallization leads to a higher temperature hot-spot, which is located within the device epitaxial layer just under the crack site. Nevertheless, these conclusions have not yet been experimentally verified and they have been achieved by using a simplified model of the crack.

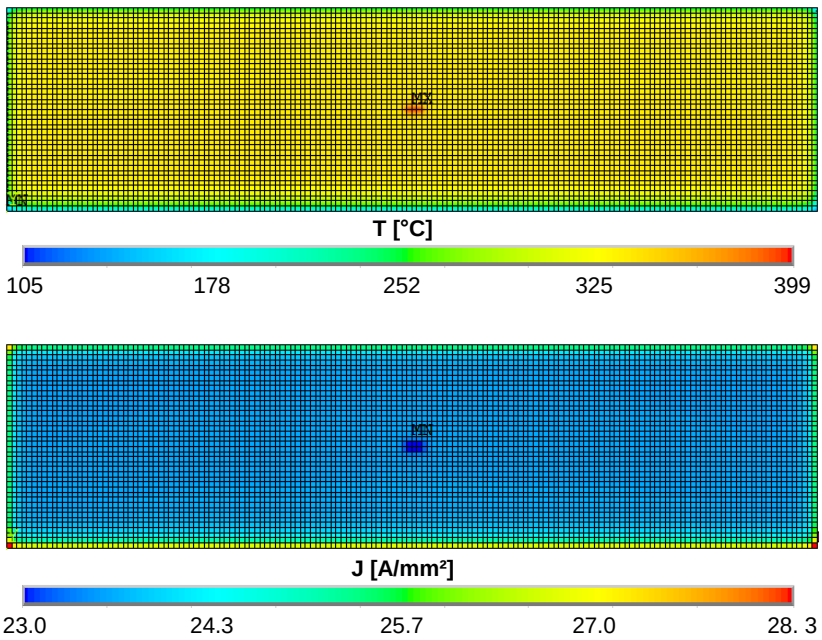

Figure 4.13: Temperature (top) and current density (bottom) distribution at the temperature peak within the device active region of the FE model featuring a 51.2 µm crack (4 x 2 void elements).

Conclusions and outlook

In the research activity reported in this dissertation an FEM-based Power MOSFETs electro-thermal simulator has been developed by customizing ANSYS® software. Two algorithms have been conceived for suitably simulating electro-thermal behavior of discrete power MOSFETs and of their integrated version within Smart Power switches. With those methods the device electro-thermal response to a single stress pulse during transient operation can be predicted. The computational effort required to perform an electro-thermal simulation may vary substantially depending on:

- the number of nodes/elements of the finite element model (representing the device geometry) and its complexity;
- the number of elements employed to mesh the active area of the device (i.e. the so defined *channel elements*);
- the type of modeling adopted for the epitaxial region: COMPLEX modeling is computationally more demanding than SIMPLE modeling due to highly anisotropic equivalent material properties and to the larger number of elements required;
- the simulation algorithm: simulations using I_{D} as an input waveform are slower because their algorithm requires more iterations;
- the number of time steps used to sample the input waveforms.

Electro-thermal simulations presented in this work required a computational time varying from about 4 to 9 hours on an 8-core workstation with hyper-threading. Computational time represents the bottleneck of the developed algorithms, although computational improvements can be eventually achieved by adopting more effective and element-saving meshing strategies, such as non-matching grids.

Simulation results of thermally stable and unstable biasing conditions agree with physical considerations. Furthermore, experimental measurements have been performed and used to validate simulation results. For this purpose, bipolar temperature sensors integrated in two different test chips have been calibrated in a limited temperature range due to intrinsic setup limitations. The measurement accuracy of extrapolated temperature values (i.e. out of the calibration range) relies on the accuracy of the analytical model (linear) adopted to fit the calibration data. The good matching obtained between measured and simulated temperatures proves the accuracy of simulation results. Nevertheless, more robust temperature sensor concepts together with the use of a wider temperature calibration range may help to better validate simulation results at higher temperatures. Furthermore, by using a test device integrating six temperature sensors within the whole DMOS area, a very good spatial match has been obtained between the simulated temperature map and measured sensor temperatures. For this kind of verification, the use of an optical method, such as infrared thermography, would be more appropriate to verify the simulated temperature distribution over the device active area.

The electro-thermal modeling of the elementary cell has been performed by adopting simple analytical descriptions already documented in literature. Those models have been calibrated by fitting temperature dependent transfer characteristics obtained by calibrated two-dimensional device simulations. Several electro-thermal simulations performed in a limited V_{GS}-range (i.e. few volts above the V_{th}) have proven that using a simple SPICE level-1 model (including temperature dependencies of the mobility and the threshold voltage) for the cell behavioral description returns simulation results that well agree with the ones obtained using the table model. Therefore, the choice of the analytical model should be taken considering the trade-off between model accuracy and model simplicity. Consequently, electro-thermal simulations are sufficiently accurate when the fitting error is small enough. Moreover, the definition of temperature dependent (nonlinear) material properties, especially for silicon, is extremely important for achieving an adequate level of simulation accuracy.

With the developed approach arbitrary device geometries can be simulated. An obvious trade-off is established between complexity of the model and computational effort. The homogenization strategy reported in the third chapter is a well-known technique within the numerical computation community. Since a power device represents a typical case of multi-scale geometry, the homogenization concept represents a very suitable solution to be applied to power

MOSFET geometries in order to overcome the mentioned trade-off. Whit this technique, it has been demonstrated that an adequate modeling of the epitaxial layer of a power trench MOSFET may become necessary to achieve accuracy when simulating high power biasing conditions in very fast transient events (below 10 µs), while a simple, and commonly adopted, isotropic modeling of the epitaxial layer can be employed for simulations of low power pulses. Nevertheless, the achieved results would need to be experimentally verified.

Simulation results reported in the fourth chapter well demonstrate the usefulness of such a simulator. The impact of the transconductance coefficient K_{MOS} on the robustness characteristic of power MOSFETs integrated in latest Smart Power switches has been assessed. Simulation results have proven, in accordance to recently published studies, that higher K_{MOS} values determine smaller current and voltage thermal stability ranges and hence a reduced robustness of the power device. The achieved results have already been predicted in literature by means of analytical considerations and experimental measurements, while here they have been demonstrated for the first time using electrothermal simulations. In the same chapter, further analyses have been conducted to estimate the impact of a crack within the copper metallization of a power MOSFET. It has been demonstrated by means of a simplified finite element modeling that such a crack causes a severe temperature hot-spot in the silicon, localized at the crack site. The larger the crack size, the higher the temperature hot spot. The accuracy of the obtained results should be validated with opportune measurements aiming at the assessment of thermal runaway occurrence. Nevertheless, first simulation analyses can be explained by means of considerations consistent with the physics background.

The proposed simulation methods are suitable for assessing the robustness of power MOSFETs in thermally stable and unstable operating conditions. It is important to underline that their employment is limited to the study of device operations within the Forward-Biased Safe Operating Area (FBSOA), i.e. for biasing conditions where the V_{DS} voltage is smaller than the avalanche breakdown voltage. Therefore, over-voltage protected devices (like most modern Smart Power switches) can be suitably simulated with the developed tool. Nevertheless, the simulation algorithm could be extended towards thermal analyses of power MOSFETs in their Reverse-Biased Safe Operating Area (RBSOA). In this case, the cell analytical model must include the relevant contribution to the overall drain current determined by the generation of electron-hole pairs in the body-drain depletion region due to the impact ionization.

A further (and possibly feasible) extension of both simulation algorithms may be implemented for assessing mechanical displacements caused by the device electro-thermal behavior, which represents an approach for electro-thermo-mechanical simulation. Thermo-mechanical simulation of power devices is based on the one-directional coupling between thermal and mechanical field, i.e. the thermal field drives mechanical displacements within the device geometry. By exploiting the ANSYS capability for FEM-based mechanical analyses, the electro-thermal simulation algorithm should be modified in a way that a mechanical simulation must be performed soon after the electro-thermal iteration loop (reported in Figure 2.9). Such an electro-thermo-mechanical simulator would represent a very useful tool for assessing and analyzing mechanical fatigue phenomena driven by temperature changes. In fact, these effects are known to be the main cause of degradation mechanisms observed in modern power MOSFETs.

In conclusion, in this dissertation it has been shown that the electro-thermal interaction in power MOSFETs is responsible for thermally unstable operation which substantially limits the device SOA and hence its robustness. Its understanding and its correct prediction by means of a reliable simulation tool become crucial during the development of new technologies, the improvement of already existing ones and/or during the design phase of the device geometry.

Appendix A

Material Properties

A.1 Thermal material properties of silicon

As already mentioned in this dissertation, the silicon thermal material properties, i.e. thermal conductivity λ_{Si} and thermal capacity $C_{\mathrm{th_{Si}}}$, are nonlinear, namely their values strongly depend on the temperature. Furthermore, λ_{Si} shows a substantial dependence on the doping concentration. The definition of correct thermal material properties for silicon is of utmost importance for achieving accuracy in FEM thermal and electro-thermal simulations because the device substrate represents the main heat extraction path for the power generated in the epitaxial layer.

The thermal capacity of a material C_{th} is given by the product of its specific heat c with its density ρ (see Equation 3.9 in Paragraph 3.2.4). The value of silicon density used in FEM simulations is:

$$\rho_{\mathrm{Si}} = 2329\ \mathrm{kg\,m^{-3}}$$

and does not significantly depend on temperature.

Values of λ_{Si} and c_{Si} used in FEM simulations of this thesis have been experimentally determined by means of a dedicated material characterization method, i.e. the so called differential scanning calorimetry. Both the thermal conductivity and the specific heat of silicon have been measured over a wide range of ambient temperatures, since FEM simulations must reliably compute the thermal field according to the typically large thermal gradients and high temperature peaks observed in power MOSFETs during critical stress conditions. Experimental values of λ_{Si} and c_{Si} have been reported in Table A.1 and plotted in Figure A.1 [98].

$\mathbf{T}$ [°C]	$\boldsymbol{\lambda_{Si}}$ [W m^{-1} K^{-1}]	$\boldsymbol{c_{Si}}$ [J kg^{-1} K^{-1}]
25	74.8	706
76	67.5	750
126	61.1	782
176	55.7	806
226	50.9	824
275	47.1	837
326	43.7	849
376	40.5	858
426	37.9	867
476	35.7	876

Table A.1: Nonlinear thermal properties of silicon [98].

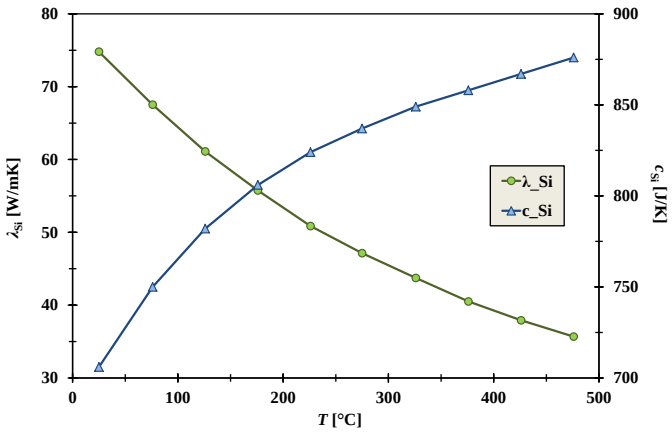

Figure A.1: Nonlinear thermal conductivity (green curve) and nonlinear specific heat (blue curve) of the highly doped silicon used in FEM simulations [98].

T [°C]	$\boldsymbol{\lambda_{\mathrm{poly}}}$ [W m^{-1} K^{-1}]
25	54.3
75	48.8
125	43.7
175	39.1
225	34.9
275	31.3
325	28.2
375	25.4
425	23.0
475	21.0

Table A.2: Nonlinear thermal conductivity of polysilicon used in FEM simulations [99].

It should be noticed that the measured silicon thermal conductivity is significantly smaller than the typical literature value of 150 W m^{-1} K^{-1} at 25°C [111, 23]. The reason for that lies in the high doping level of the bulk silicon needed to achieve very low ON-state resistance in power MOSFETs.

A.2 Thermal material properties of polysilicon

The temperature dependence of the polysilicon thermal conductivity λ_{poly} has been assessed by means of the polynomial model proposed by Lott et al. [99]. Polysilicon thermal conductivity values have been reported for different ambient temperatures in Table A.2 and plotted in Figure A.2.

The density of polysilicon used in FEM simulations is given by [99]:

$$\rho_{\mathrm{poly}} = 2330 \text{ kg m}^{-3}$$

The specific heat of polysilicon has been assumed to be equal to the silicon nonlinear specific heat (reported in Table A.1).

A.3 Thermal material properties of oxides

Two kind of oxides have been used in the FEM simulations presented in this thesis, namely silicon dioxides $\mathrm{SiO_2}$ and BPSG. For both oxides, equal thermal material properties have been assumed. Their temperature dependent thermal

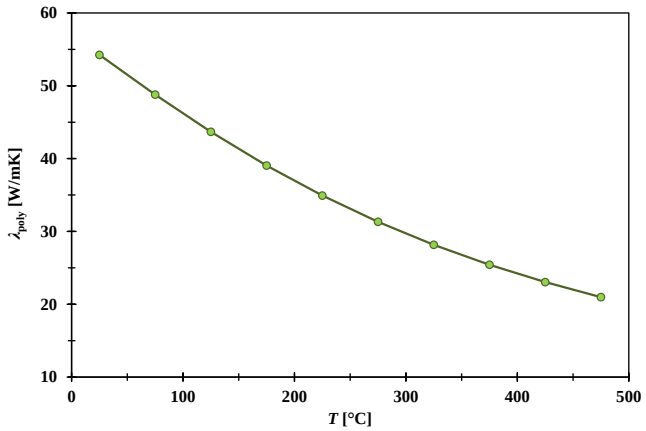

Figure A.2: Nonlinear thermal conductivity of polysilicon used in FEM simulations [99]

conductivity has been reported in Table A.3 and plotted in Figure A.3. Those values well agree with literature data [112, 113, 114, 115].

Density and specific heat of both silicon dioxides and BPSG have been assigned according to [95]:

$$\rho_{\mathrm{SiO_2}} = \rho_{\mathrm{BPSG}} = 2200\ \mathrm{kg\,m^{-3}}$$
$$c_{\mathrm{SiO_2}} = c_{\mathrm{BPSG}} = 1000\ \mathrm{J\,K^{-1}}$$

T [°C]	$\boldsymbol{\lambda_{\mathbf{SiO_2}}}$ [$\mathrm{W\,m^{-1}\,K^{-1}}$]
27	1.38
127	1.51
227	1.62
327	1.75
427	1.92
527	2.17

Table A.3: Nonlinear thermal conductivity of silicon dioxide (used also for the BPSG).

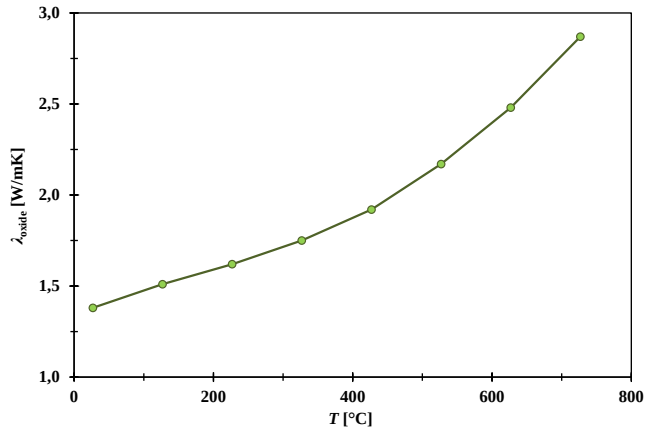

Figure A.3: Nonlinear thermal conductivity of both silicon dioxide and BPSG.

A.4 Thermal material properties of copper

Test devices (DUT1 and DUT2) employed in this thesis feature a copper-based power metallization. Generally, thermal material properties of metals do not significantly change for a wide range of temperatures, therefore only one value has been assigned for the thermal conductivity, density and specific heat of copper [115]:

$$\lambda_{\mathrm{Cu}} = 385\ \mathrm{W\,m^{-1}\,K^{-1}}$$
$$\rho_{\mathrm{Cu}} = 8930\ \mathrm{kg\,m^{-3}}$$
$$c_{\mathrm{Cu}} = 385\ \mathrm{J\,K^{-1}}$$

References

[1] B. J. Baliga, "The future of power semiconductor device technology," *Proceedings of the IEEE*, vol. 89, no. 6, pp. 822–832, 2001.

[2] R. Kraus and H. J. Mattausch, "Status and trends of power semiconductor device models for circuit simulation," *IEEE Transactions on Power Electronics*, vol. 13, no. 3, pp. 452–465, 1998.

[3] B. K. Bose, "Evaluation of modern power semiconductor devices and future trends of converters," *IEEE Transactions on Industry Applications*, vol. 28, no. 2, pp. 403–413, 1992.

[4] B. Murari, F. Bertotti, and G. A. Vignola, *Smart Power ICs: technologies and applications*. Springer, 2002, vol. 6.

[5] W. Kanert, "Reliability challenges for power devices under active cycling," in *IEEE International Reliability Physics Symposium*, 2009, pp. 409–415.

[6] W. Kanert, "Active cycling reliability of power devices: expectations and limitations," *Microelectronics Reliability*, vol. 52, no. 9-10, pp. 2336–2341, 2012.

[7] P. Alpern, P. Nelle, E. Barti, H. Gunther, A. Kessler, R. Tilgner, and M. Stecher, "On the way to zero defect of plastic-encapsulated electronic power devices—part I: metallization," *IEEE Transactions on Device and Materials Reliability*, vol. 9, no. 2, pp. 269–278, 2009.

[8] P. Alpern, P. Nelle, E. Barti, H. Gunther, A. Kessler, R. Tilgner, and M. Stecher, "On the way to zero defect of plastic-encapsulated electronic power devices—part II: molding compound," *IEEE Transactions on Device and Materials Reliability*, vol. 9, no. 2, pp. 279–287, 2009.

[9] P. Alpern, P. Nelle, E. Barti, H. Gunther, A. Kessler, R. Tilgner, and M. Stecher, "On the way to zero defect of plastic-encapsulated electronic power devices—part III: chip coating, passivation and design," *IEEE Transactions on Device and Materials Reliability*, vol. 9, no. 2, pp. 288–295, 2009.

[10] M. Nelhiebel, R. Illing, C. Schreiber, S. Wöhlert, S. Lanzerstorfer, M. Ladurner, C. Kadow, S. Decker, D. Dibra, H. Unterwalcher *et al.*, "A reliable technology concept for active power cycling to extreme temperatures," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1927–1932, 2011.

[11] T. Smorodin, J. Wilde, P. Alpern, and M. Stecher, "A temperature-gradient-induced failure mechanism in metallization under fast thermal cycling," *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 3, pp. 590–599, 2008.

[12] P. Spirito, G. Breglio, V. d'Alessandro, and N. Rinaldi, "Thermal instabilities in high current power MOS devices: experimental evidence, electro-thermal simulations and analytical modeling," in *IEEE 23rd International Conference on Microelectronics (MIEL)*, vol. 1, 2002, pp. 23–30.

[13] A. Consoli, F. Gennaro, A. Testa, G. Consentino, F. Frisina, R. Letor, and A. Magri, "Thermal instability of low voltage power-MOSFETs," *IEEE Transactions on Power Electronics*, vol. 15, no. 3, pp. 575–581, 2000.

[14] H. Köck, R. Illing, T. Ostermann, S. Decker, D. Dibra, G. Pobegen, S. de Filippis, M. Glavanovics, and D. Pogany, "Design of a test chip with small embedded temperature sensor structures realized in a common-drain power trench technology," in *IEEE International Conference on Microelectronic Test Structures (ICMTS)*, 2011, pp. 176–181.

[15] M. Pfost, D. Costachescu, A. Podgaynaya, M. Stecher, S. Bychikhin, D. Pogany, and E. Gornik, "Small embedded sensors for accurate temperature measurements in DMOS power transistors," in *IEEE International Conference on Microelectronic Test Structures (ICMTS)*, 2010, pp. 3–7.

[16] D. Dibra, C. Kadow, M. Pfost, N. Krischke, A. Lindemann, J. Lutz, and M. Stecher, "Scaling of temperature sensors for Smart Power MOSFETs," *International Seminar on Power Semiconductors (ISPS)*, pp. 139–148, 2008.

[17] A. Irace, "Infrared thermography application to functional and failure analysis of electron devices and circuits," *Microelectronics Reliability*, vol. 52, no. 9-10, pp. 2019–2023, 2012.

[18] S. de Filippis, "Development of an experimental setup for infrared camera characterization aimed at Smart Power devices thermography," Master's thesis, Università degli Studi di Napoli "Federico II" - Dipartimento di Ingegneria Biomedica, Elettronica e delle Telecomunicazioni, 2009.

[19] H. Köck, S. de Filippis, C. Djelassi, H.-P. Kreuter, R. Illing, M. Glavanovics, and M. Kaltenbacher, "In-situ calibration procedure for infrared microscopy temperature measurements," *Proc. of the Microtechnology and Thermal Problems in Electronics*, pp. 102–107, 2011.

[20] G. Haberfehlner, S. Bychikhin, V. Dubec, M. Heer, A. Podgaynaya, M. Pfost, M. Stecher, E. Gornik, and D. Pogany, "Thermal imaging of Smart Power DMOS transistors in the thermally unstable regime using a compact transient interferometric mapping system," *Microelectronics Reliability*, vol. 49, no. 9, pp. 1346–1351, 2009.

[21] D. Pogany, S. Bychikhin, C. Furbock, M. Litzenberger, E. Gornik, G. Groos, K. Esmark, and M. Stecher, "Quantitative internal thermal energy mapping of semiconductor devices under short current stress using backside laser interferometry," *IEEE Transactions on Electron Devices*, vol. 49, no. 11, pp. 2070–2079, 2002.

[22] H. Köck, "Experimental and numerical study on heat transfer problems in microelectronic devices," Ph.D. dissertation, Alpen-Adria Universität Klagenfurt, Fakultät für Technische Wissenschaften, 2012.

[23] V. Košel, R. Sleik, and M. Glavanovics, "Transient non-linear thermal FEM simulation of Smart Power switches and verification by measurements," in *IEEE 13th International Workshop on Thermal Investigation of ICs and Systems (THERMINIC)*, 2007, pp. 110–114.

[24] O. Alatise, I. Kennedy, G. Petkos, K. Khan, A. Koh, and P. Rutter, "Understanding linear-mode robustness in low-voltage trench power MOSFETs," *IEEE Transactions on Device and Materials Reliability*, vol. 10, no. 1, pp. 123–129, 2010.

[25] D. Dibra, M. Stecher, S. Decker, A. Lindemann, J. Lutz, and C. Kadow, "On the origin of thermal runaway in a trench power MOSFET," *IEEE Transactions on Electron Devices*, vol. 58, no. 10, pp. 3477–3484, 2011.

[26] S. Wünsche, C. Clauß, P. Schwarz, and F. Winkler, "Electro-thermal circuit simulation using simulator coupling," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 5, no. 3, pp. 277–282, 1997.

[27] W. Van Petegem, B. Geeraerts, W. Sansen, and B. Graindourze, "Electrothermal simulation and design of integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 2, pp. 143–146, 1994.

[28] M. Vellvehi, X. Jordà, P. Godignon, C. Ferrer, and J. Millán, "Coupled electro-thermal simulation of a DC/DC converter," *Microelectronics Reliability*, vol. 47, no. 12, pp. 2114–2121, 2007.

[29] M. Riccio, G. De Falco, L. Maresca, G. Breglio, E. Napoli, A. Irace, Y. Iwahashi, and P. Spirito, "3D electro-thermal simulations of wide area power devices operating in avalanche condition," *Microelectronics Reliability*, 2012.

[30] A. Irace, G. Breglio, and P. Spirito, "New developments of THERMOS3, a tool for 3D electro-thermal simulation of Smart Power MOSFETs," *Microelectronics Reliability*, vol. 47, no. 9, pp. 1696–1700, 2007.

[31] M. Pfost, J. Joos, and M. Stecher, "Measurement and simulation of self-heating in DMOS transistors up to very high temperatures," in *IEEE 20th International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, 2008, pp. 209–212.

[32] P. Cova, M. Bernardoni, N. Delmonte, and R. Menozzi, "Dynamic electro-thermal modeling for power device assemblies," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1948–1953, 2011.

[33] T. Hauck, W. Teulings, and E. Rudnyi, "Electro-thermal simulation of multi-channel power devices on PCB with SPICE," in *IEEE 15th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC)*, 2009, pp. 124–129.

[34] S. de Filippis, V. Košel, D. Dibra, S. Decker, H. Köck, and A. Irace, "ANSYS based 3-D electro-thermal simulations for the evaluation of power MOSFETs robustness," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1954–1958, 2011.

[35] V. Košel, S. de Filippis, L. Chen, S. Decker, and A. Irace, "FEM simulation approach to investigate electro-thermal behavior of power transistors in 3-D," *Microelectronics Reliability*, vol. 53, no. 3, pp. 356–362, 2013.

[36] S. de Filippis, R. Illing, M. Nelhiebel, S. Decker, H. Köck, and A. Irace, "Validated electro-thermal simulations of two different power MOSFET technologies and implications on their robustness," in *IEEE 23rd International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, 2013 (accepted).

[37] M. G. Geers, V. Kouznetsova, and W. Brekelmans, "Multi-scale computational homogenization: trends and challenges," *Journal of Computational and Applied Mathematics*, vol. 234, no. 7, pp. 2175–2182, 2010.

[38] A. Y. Beliaev and S. Kozlov, "Darcy equation for random porous media," *Communications on pure and applied mathematics*, vol. 49, no. 1, pp. 1–34, 1996.

[39] D. Bigaud, J.-M. Goyheneche, and P. Hamelin, "A global-local nonlinear modelling of effective thermal conductivity tensor of textile-reinforced composites," *Composites Part A: Applied Science and Manufacturing*, vol. 32, no. 10, pp. 1443–1453, 2001.

[40] I. Özdemir, W. Brekelmans, and M. Geers, "Computational homogenization for heat conduction in heterogeneous solids," *International journal for numerical methods in engineering*, vol. 73, no. 2, pp. 185–204, 2008.

[41] S. de Filippis, H. Köck, M. Nelhiebel, V. Košel, S. Decker, M. Glavanovics, and A. Irace, "Modeling of highly anisotropic microstructures

for electro-thermal simulations of power semiconductor devices," *Microelectronics Reliability*, vol. 52, no. 9-10, pp. 2374–2379, 2012.

[42] M. H. Rashid, *Power electronics handbook*. Academic Press, 2001.

[43] S. Sze and K. Ng, *Physics of semiconductor devices*. Wiley-interscience, 2006.

[44] M. Denison, "Single stress safe operating area of DMOS transistors integrated in Smart Power technologies," Ph.D. dissertation, University of Bremen, Faculty of Physics, Electrical Engineering and Information Technology, 2004.

[45] H. Keller, "Robustness validation - An improved qualification method for semiconductor devices in automotive," in *VDE 5th International Conference on Integrated Power Systems (CIPS)*, 2008, pp. 1–5.

[46] Z. E. Components and Systems, *Handbook for Robustness Validation of Semiconductor Devices in Automotive Applications*. SAE International, 2007.

[47] O. Bluder, M. Glavanovics, and J. Pilz, "Applying Bayesian mixtures-of-experts models to statistical description of Smart Power semiconductor reliability," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1464 –1468, 2011.

[48] O. Bluder, "Prediction of smart power device lifetime based on bayesian modeling," Ph.D. dissertation, Alpen-Adria University of Klagenfurt, 2011.

[49] M. Glavanovics, H. Köck, V. Košel, and T. Smorodin, "Flexible active cycle stress testing of Smart Power switches," *Microelectronics Reliability*, vol. 47, no. 9, pp. 1790–1794, 2007.

[50] H. Glavanovics, M. Köck, H. Eder, , V. Košel, and T. Smorodin, "A new cycle test system emulating inductive switching waveforms," in *IEEE 11th European Conference on Power Electronics and Applications (EPE)*, 2007, pp. 1–9.

[51] M. Glavanovics, H.-P. Kreuter, R. Sleik, and C. Schreiber, "Cycle stress test equipment for automated short circuit testing of Smart Power

switches according to the AEC Q100-012 standard," in *IEEE 13th European Conference on Power Electronics and Applications (EPE)*, 2009, pp. 1–7.

[52] M. Glavanovics, R. Sleik, and C. Schreiber, "A compact high current system for short circuit testing of Smart Power switches according to AEC standard Q100-012," in *IEEE Power Electronics Specialists Conference (PESC)*, 2008, pp. 1828–1832.

[53] N. Seliger, E. Wolfgang, G. Lefranc, H. Berg, and T. Licht, "Reliable power electronics for automotive applications," *Microelectronics Reliability*, vol. 42, no. 9-11, pp. 1597–1604, 2002.

[54] B. Baliga, *Fundamentals of Power Semiconductor Devices*. Springer, 2008.

[55] D. P. Foty, *MOSFET modeling with SPICE: principles and practice*. Prentice-Hall, Inc., 1997.

[56] Y. Chen, X. Cheng, Y. Liu, Y. Fu, T. Wu, and Z. Shen, "Modeling and analysis of metal interconnect resistance of power MOSFETs with ultra low on-resistance," in *IEEE International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, 2006, pp. 1–4.

[57] H. Köck, C. Djelassi, S. de Filippis, R. Illing, M. Nelhiebel, M. Ladurner, M. Glavanovics, and D. Pogany, "Improved thermal management of low voltage power devices with optimized bond wire positions," *Microelectronics Reliability*, vol. 51, no. 9, pp. 1913–1918, 2011.

[58] J. Sauveplane, P. Tounsi, E. Scheid, and A. Deram, "3D electro-thermal investigations for reliability of ultra low ON state resistance power MOSFET," *Microelectronics Reliability*, vol. 48, no. 8, pp. 1464–1467, 2008.

[59] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog integrated circuits and signal processing*, vol. 8, no. 1, pp. 83–114, 1995.

[60] M. Pfost, D. Costachescu, A. Podgaynaya, and M. Stecher, "A simple approach for DMOS transistor modeling up to very high temperatures," in *IEEE International Semiconductor Conference (CAS)*, vol. 2, 2009, pp. 389–392.

[61] V. d'Alessandro, "Analysis of electro-thermal effects in semiconductor devices," Ph.D. dissertation, Università degli Studi di Napoli "Federico II", Dipartimento di Ingengeria Elettronica e delle Telecomunicazioni, 2002.

[62] M. Darwish, "Study of the quasi-saturation effect in VDMOS transistors," *IEEE Transactions on Electron Devices*, vol. 33, no. 11, pp. 1710–1716, 1986.

[63] C. Kreuzer, N. Krischke, and P. Nance, "Physically based description of quasi-saturation region of vertical DMOS power transistors," in *IEEE International Electron Devices Meeting (IEDM)*, 1996, pp. 489–492.

[64] K. Hoffmann, *System Integration: from transistor design to large scale integrated circuits*. Wiley, 2004.

[65] N. Arora, *MOSFET modeling for VLSI simulation: theory and practice*. World Scientific Publishing Company Incorporated, 2007.

[66] P. Spirito, G. Breglio, V. d'Alessandro, and N. Rinaldi, "Analytical model for thermal instability of low voltage power MOS and SOA in pulse operation," in *IEEE 14th International Symposium on Power Semiconductor Devices and ICs (ISPDS)*, 2002, pp. 269–272.

[67] P. Spirito, G. Breglio, and V. d'Alessandro, "Modeling the onset of thermal instability in low voltage power MOS: an experimental validation," in *IEEE 17th International Symposium on Power Semiconductor Devices and ICs (ISPSD)*, 2005, pp. 183–186.

[68] P. Hover and G. Govil, *IEEE Transactions on Electron Devices*, vol. 21, pp. 617–623, 1974.

[69] G. Breglio, N. Rinaldi, and P. Spirito, "Thermal mapping and 3D numerical simulation of new cellular power MOS affected by electro-thermal instability," *Microelectronics Journal*, vol. 31, no. 9, pp. 741–746, 2000.

[70] C. Kadow, S. Decker, D. Dibra, N. Krischke, S. Lanzerstorfer, H. Maier, T. Meyer, N. Vannucci, and R. Zink, "Fabrication of trench isolation and trench power MOSFETs in a Smart Power IC technology with a single trench unit process," in *IEEE 21st International Symposium on Power Semiconductor Devices and IC's (ISPSD)*, 2009, pp. 224–226.

[71] P. A. Basore, "Numerical modeling of textured silicon solar cells using PC-1D," *IEEE Transactions on Electron Devices*, vol. 37, no. 2, pp. 337–343, 1990.

[72] D. A. Clugston and P. A. Basore, "PC1D version 5: 32-bit solar cell modeling on personal computers," in *IEEE 26th Photovoltaic Specialists Conference*, 1997, pp. 207–210.

[73] M. Ladurner, R. Illing, P. Del Croce, and B. Auer, "Power switch temperature control device and method," Patent, Jun. 28, 2010, US Patent App. 12/824,891.

[74] H. Köck, R. de Filippis, S. ans Sleik, M. Glavanovics, and D. Pogany, "A transient temperature mapping system for integrated circuits operating in the microsecond time scale," in *18th Workshop on Microelectronics (AUSTROCHIP)*, 2010, pp. 73–78.

[75] Datasheet of Infineon PROFET BTS5020-2EKA.

[76] H. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*. Oxford Science Publications, 1946.

[77] D. Fleisch, *A Student's Guide to Maxwell's Equations*. Cambridge University Press, 2009.

[78] S.-M. Lee and D. G. Cahill, "Heat transport in thin dielectric films," *Journal of Applied Physics*, vol. 81, no. 6, pp. 2590–2595, 1997.

[79] M. Ciappa, W. Fichtner, T. Kojima, Y. Yamada, and Y. Nishibe, "Extraction of accurate thermal compact models for fast electro-thermal simulation of IGBT modules in hybrid electric vehicles," *Microelectronics Reliability*, vol. 45, no. 9, pp. 1694–1699, 2005.

[80] V. Košel, "Thermal impedance modelling of Smart Power switches," Master's thesis, Slovak University of Technology Bratislava, Faculty of Electrical Engineering and Information Technology, Department of Microelectronics, 2003.

[81] E. Rudnyi and J. Korvink, "Model order reduction for large scale engineering models developed in ANSYS," *Applied Parallel Computing, State of the Art in Scientific Computing*, pp. 349–356, 2006.

[82] M. Kaltenbacher, *Numerical simulation of mechatronic sensors and actuators*. Springer Verlag, 2004.

[83] H. Köck, V. Košel, C. Djelassi, M. Glavanovics, and D. Pogany, "IR thermography and FEM simulation analysis of on-chip temperature during thermal-cycling power-metal reliability testing using in situ heated structures," *Microelectronics Reliability*, vol. 49, no. 9, pp. 1132–1136, 2009.

[84] R. Lewis, K. Morgan, H. Thomas, and K. Seetharamu, *The finite element method in heat transfer analysis*. Wiley, 1996.

[85] T. Hughes, *The finite element method: linear static and dynamic finite element analysis*. Dover Publications, 2000.

[86] O. Zienkiewicz, R. Taylor, and J. Zhu, *The finite element method: its basis and fundamentals*. Butterworth-Heinemann, 2005, vol. 1.

[87] E. Antonova and D. Looman, "Finite elements for thermoelectric device analysis in ANSYS," in *IEEE 24th International Conference on Thermoelectrics (ICT)*, 2005, pp. 215–218.

[88] *ANSYS Mechanical APDL and Mechanical Applications, Theory Reference - Release 13.0*, 2010.

[89] G. Pobegen, M. Nelhiebel, S. de Filippis, and T. Grasser, "Accurate high temperature measurements using local polysilicon heater structures," *IEEE Transactions on Device and Materials Reliability*, no. 1, 2013 (submitted for publication).

[90] V. Košel, C. Djelassi, H. Köck, M. Glavanovics, and A. Zechmann, "Electro-thermal simulation of a semiconductor device based on simulatively extracted electrical parameters from measurements," in *Proc. of EUROSIM Congress*, 2010.

[91] C. Djelassi, "Investigation on degradation of aluminum lines caused by fast and defined temperature cycling," Master's thesis, Alpen-Adria Universität Klagenfurt, Fakultät für Technische Wissenschaften, 2011.

[92] E. Madenci and I. Guven, *The finite element method and applications in engineering using ANSYS*. Springer, 2007.

[93] S. Reggiani, M. Valdinoci, L. Colalongo, M. Rudan, G. Baccarani, A. Stricker, F. Illien, N. Felber, W. Fichtner, and L. Zullino, "Electron and hole mobility in silicon at large operating temperatures – part I. Bulk mobility," *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 490–499, 2002.

[94] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[95] S. Franssila, *Introduction to microfabrication*. Wiley, 2010.

[96] A. Castellazzi, V. Kartal, R. Kraus, N. Seliger, M. Honsberg-Riedl, and D. Schmitt-Landsiedel, "Hot-spot meaurements and analysis of electro-thermal effects in low-voltage power-MOSFET's," *Microelectronics Reliability*, vol. 43, no. 9-11, pp. 1877–1882, 2003.

[97] H. Köck, S. de Filippis, M. Nelhiebel, M. Glavanovics, and M. Kaltenbacher, "Multiscale FE modeling concepts applied to micro-electronic device simulations," in *IEEE 14th International Conference on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro/Nano-Electronics and Microsystems (EuroSimE)*. IEEE, 2013 (accepted).

[98] H.-P. Kreuter, "Virtual current and temperature sensors for Smart Power switches," Ph.D. dissertation, Vienna University of Technology, 2013.

[99] C. Lott, T. McLain, J. Harb, and L. Howell, "Thermal modeling of a surface-micromachined linear thermomechanical actuator," 2001, pp. 370–373.

[100] S. Eiser, M. Kaltenbacher, and M. Nelhiebel, "Non-conforming meshes in multi-scale thermo-mechanical Finite Element analysis of semiconductor power devices," in *IEEE International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Micro-electronics and Microsystems (EuroSimE)*, 2013 (accepted).

[101] C. Bernardi, "A new nonconforming approach to domain decomposition: the mortar element method," *Nonliner Partial Differential Equations and Their Applications*, 1994.

[102] M. N. Özisik and H. R. Orlande, *Inverse heat transfer*. Taylor & Francis New York, 2000.

[103] S. T. Peake, R. Grover, R. Farr, C. Rogers, and G. Petkos, "Fully self-aligned power trench-MOSFET utilising 1 μm pitch and 0.2 μm trench width," in *IEEE 14th International Symposium onPower Semiconductor Devices and ICs (ISPSD)*, pp. 29–32, 2002.

[104] F.-T. Chien, C.-N. Liao, C.-L. Wang, H.-C. Chiu, and Y.-T. Tsai, "Low on-resistance trench power MOSFETs design," *IET Electronics Letters*, vol. 44, no. 3, pp. 232–234, 2008.

[105] K. Shenai, "Optimized trench MOSFET technologies for power devices," *IEEE Transactions on Electron Devices*, vol. 39, no. 6, pp. 1435–1443, 1992.

[106] P. Goarin, R. van Dalen, G. Koops, and C. Le Cam, "Power trench MOSFETs with very low specific on-resistance for 25V applications," *Solid-State Electronics*, vol. 51, no. 11, pp. 1589–1595, 2007.

[107] M. Glavanovics and H. Zitta, "Thermal destruction testing: an indirect approach to a simple dynamic thermal model of Smart Power switches," in *IEEE 27th European Solid-State Circuits Conference (ESSCIRC)*, 2001, pp. 221–224.

[108] T. Detzel, M. Glavanovics, and K. Weber, "Analysis of wire bond and metallization degradation mechanisms in DMOS power transistors stressed under thermal overload conditions," *Microelectronics Reliability*, vol. 44, no. 9-11, pp. 1485–1490, 2004.

[109] S. Russo, R. Letor, O. Viscuso, L. Torrisi, and G. Vitali, "Fast thermal fatigue on top metal layer of power devices," *Microelectronics Reliability*, vol. 42, no. 9-11, pp. 1617–1622, 2002.

[110] S. Gopalan, B. Krabbenborg, J.-H. Egbers, B. van Velzen, and R. Zingg, "Reliability of power transistors against application driven temperature swings," *Microelectronics Reliability*, vol. 42, no. 9-11, pp. 1623–1628, 2002.

[111] C. Glassbrenner and G. A. Slack, "Thermal conductivity of silicon and germanium from 3 K to the melting point," *Physical Review*, vol. 134, no. 4A, pp. A1058–A1069, 1964.

[112] H. A. Schafft, J. S. Suehle, and P. Mirel, "Thermal conductivity measurements of thin-film silicon dioxide," in *IEEE International Conference on Microelectronic Test Structures (ICMTS)*, 1989, pp. 121–125.

[113] J. R. Black, "Current limitations of thin film conductors," in *IEEE 20th Annual Reliability Physics Symposium*, 1982, pp. 300–306.

[114] J. Lee and B. Cho, "Large time-scale electro-thermal simulation for loss and thermal management of power MOSFET," in *IEEE 34th Annual Power Electronics Specialist Conference (PESC)*, vol. 1, 2003, pp. 112–117.

[115] J. F. Shackelford and W. Alexander, *CRC materials science and engineering handbook*. CRC press, 2010.

Acronyms

APDL	ANSYS® Parametric Design Language
BC	Boundary Condition
BJT	Bipolar Junction Transistor
BPSG	Borophosphosilicate Glass
CAE	Computer-Aided Engineering
CTE	Coefficient of Thermal Expansion
DMOS	Diffused MOSFET
DUT	Device Under Test
ESD	Electrostatic Discharge
FBSOA	Forward-Biased Safe Operating Area
FE	Finite Element
FEM	Finite Element Method
FIB	Focused Ion Beam
IC	Integrated Circuit
IGBT	Insulated-Gate Bipolar Transistor
ILD	Inter-Layer Dielectric
JFET	Junction Field Effect Transistor

MOSFET	Metal–Oxide–Semiconductor Field Effect Transistor
PDE	Partial Differential Equation
PWM	Pulse Width Modulation
RBSOA	Reverse-Biased Safe Operating Area
SC	Short Circuit
SEM	Scanning Electron Microscope
SOA	Safe Operating Area
SP	Smart Power
TCP	Temperature Compensation Point
VD-MOSFET	Vertical-Diffused MOSFET

Symbols

Symbol	Description	Unit
A	Area	m^2
C_{th}	Thermal capacity	$\mathrm{J\,K}^{-1}$
D_{n}	Electron diffusivity	$\mathrm{cm}^2\,\mathrm{s}^{-1}$
D_{p}	Hole diffusivity	$\mathrm{cm}^2\,\mathrm{s}^{-1}$
E_{g}	Band gap energy	eV
I_{D}	Drain current	A
I_{TCP}	Drain current level for $\alpha = 0$	A
J	Current density	$\mathrm{A\,m}^{-2}$
K_{MOS}	Transconductance coefficient, or conductivity parameter, of a power MOSFET	$\mathrm{A\,V}^{-2}$
L_{ch}	Channel length of a power MOSFET	m
M	Exponent of the temperature dependent mobility	-
N	Basis function, or shape function, for the FE formulation	-
N_{A}	Acceptor concentration	cm^{-3}
N_{D}	Donor concentration	cm^{-3}
Q	Heat generation term	J
Q_{V}	Volumetric heat generation term	$\mathrm{W\,m}^{-3}$
Q_{el}	Joule heat dissipation term	$\mathrm{W\,m}^{-3}$
R_{B}	Base resistance of a generic BJT	Ω
R_{ON}	ON-state resistance of a power device	Ω
R_{ch}	Equivalent resistance associated to the channel of a power MOSFET	Ω
R_{drift}	Equivalent resistance associated to the drift region of a power MOSFET	Ω
R_{s}	Source resistance of a power MOSFET	Ω
T	Absolute temperature	K

Symbol	Description	Unit
$T(\boldsymbol{x}, t)$	Temperature field	K
T_0	Reference temperature	K
T_{HS}	Temperature hot-spot	K
T_{c}	Case temperature	K
T_{j}	Junction temperature	K
V	Volume	m^3
V_{BE}	Base-emitter voltage of a generic BJT	V
V_{DS}	Drain-source voltage	V
V_{D}	Drain potential	V
V_{GS}	Gate-source voltage	V
V_{G}	Gate potential	V
V_{S}	Source potential	V
V_{TCP}	Gate-source voltage for $\alpha = 0$	V
V_{T}	Thermal voltage	V
V_{th}	Threshold voltage	V
V_{th_0}	Threshold voltage at the reference temperature T_0	V
W_{cell}	Channel width of the elementary cell in a power MOSFET	m
W_{eff}	Effective channel width of a power MOSFET	m
$Z_{\text{th}_{\text{j-c}}}$	Junction to case thermal impedance	$\text{K}\,\text{W}^{-1}$
Γ	Domain boundary	m^2
Ω	Space domain	m^3
Φ	Threshold voltage temperature coefficient	$\text{V}\,\text{K}^{-1}$
α	Drain current temperature coefficient	$\text{A}\,\text{s}^{-1}$
β	Common-emitter current gain of a generic BJT	-
$\boldsymbol{B}$	Magnetic flux density	T
$\boldsymbol{E}$	Electric field	$\text{V}\,\text{m}^{-1}$
$\boldsymbol{J}$	Current density vector	$\text{A}\,\text{m}^{-2}$
$\boldsymbol{f}_T$	Right-hand side, or load vector, of the thermal FE formulation	-
$\boldsymbol{f}_\phi$	Right-hand side, or load vector, of the electrical FE formulation	-
$\boldsymbol{q}$	Heat flux	$\text{W}\,\text{m}^{-2}$
ϵ_{OX}	Absolute permittivity of the gate oxide in a power MOSFET	$\text{F}\,\text{m}^{-1}$
ϵ_{s}	Absolute permittivity of silicon	$\text{F}\,\text{m}^{-1}$

Symbol	Description	Unit
η	Generic direction in a Cartesian coordinate system	m
$\hat{\mathbf{n}}$	Outward versor normal to the domain boundary Γ	m
λ	Thermal conductivity	$\mathrm{W\,m^{-1}\,K^{-1}}$
$\mathbf{C}_T$	Capacity matrix, or thermal damping matrix, of the FE formulation	-
$\mathbf{K}_T$	Conductivity matrix, or thermal stiffness matrix, of the FE formulation	-
$\mathbf{K}_\phi$	Electric stiffness matrix of the FE formulation	-
μ_0	Electron mobility at the reference temperature T_0	$\mathrm{cm^2\,V^{-1}\,s^{-1}}$
μ_n	Electron mobility	$\mathrm{cm^2\,V^{-1}\,s^{-1}}$
ω	Test function for the FE formulation	-
ϕ	Electric scalar potential	V
ρ	Electrical resistivity	$\Omega\,\mathrm{m^{-1}}$
ρ	Mass density (per unit volume)	$\mathrm{kg\,m^{-3}}$
ρ_e	Density of charge per unit volume	$\mathrm{C\,m^{-3}}$
σ	Electrical conductivity	S
τ_g	Carrier generation lifetime	µs
τ_n	Electron lifetime	µs
τ_p	Hole lifetime	µs
k	Boltzmann constant k = $1.380\,65 \times 10^{-23}$	$\mathrm{J\,K^{-1}}$
q	Elementary charge q = $1.620\,21 \times 10^{-19}$	C
$\tilde{\lambda}$	Thermal conductivity tensor	$\mathrm{W\,m^{-1}\,K^{-1}}$
$\tilde{\sigma}$	Electrical conductivity tensor	S
c	Specific heat	$\mathrm{J\,kg^{-1}\,K^{-1}}$
d	Thickness	m
g_m	Differential transconductance parameter	S
h	Heat transfer coefficient	$\mathrm{W\,m^{-2}\,K^{-1}}$
m	Mass	kg
n_i	Intrinsic carrier concentration in silicon	$\mathrm{cm^{-3}}$
q_n	Outward normal heat flux	$\mathrm{W\,m^{-2}}$
t	Time	s
t_OX	Thickness of the gate oxide in a power MOSFET	m